

FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO
UNIVERZA V LJUBLJANI

Luka Fürst

Izbira značilnic za detekcijo objektov
dane vizualne kategorije na slikah

MAGISTRSKA NALOGA

Mentor: prof. dr. Aleš Leonardis

Ljubljana, 2007

Zahvala

Prof. dr. Aleš Leonardis me je vodil skozi zahtevna, izzivov polna, občasno tudi nekoliko naporna, vendar po svoje neponovljiva leta magistrskega študija. Prof. Leonardis me je naučil, kako pristopiti k raziskovalnemu delu, kako opravljeno delo pošteno ovrednotiti, kako dobljene rezultate predstaviti v objavljenem članku, kako se lotiti znanstvenih recenzij, pa še marsikaj drugega sem po njegovi zaslugi spoznal v študijskem obdobju. Na primer to, da si znanstvene avtoritete sicer zaslužijo spoštovanje, kljub temu pa kritična in zdravorazumska presoja nikdar ni odveč. Pa tudi to, da za dosego znanstvenega uspeha ni dovolj, če raziskovanje obravnavamo kot običajno (rutinsko) službo. Za vse, kar me je (na)učil, se prof. Leonardisu iskreno zahvaljujem.

Poleg prof. Leonardisa sta me v znanstvenoraziskovalno delo uvajala tudi Sanja Fidler in doc. dr. Danijel Skočaj, za kar se obema zahvaljujem. Gdč. Fidler se zahvaljujem tudi za pomoč pri pridobivanju učnih slikovnih zbirk, uporabljenih v magistrski nalogi. Prispevek gdč. Fidler in doc. Skočaja k magistrski nalogi je dejansko mnogo večji, kot bi bilo mogoče sklepati na osnovi tega skromnega odstavka.

Gregor Berginc me je navdušil za operacijski sistem Linux in urejevalnik Vim, ki sta me spremljala skozi celoten magistrski študij, vključno z izdelavo magistrske naloge. Vendar pa me Gregor ni zgolj uvedel v delo z Linuxom in Vimom, ampak mi je tudi brez obotavljanja priskočil na pomoč, kadar se je moje poznavanje enega ali drugega izkazalo za prešibko. Enako velja za drugega laboratorijskega mojstra za Linux, Aleša Štimca. Tudi njemu se zahvaljujem za izkazano pomoč.

„Beseda tuja, beseda domača, vselej vprašaj dr. Artača!“ sem nezavedno vzkliknil, kadarkoli je dr. Matej Artač razrešil kakšno perečo jezikoslovno zagato. K moji lingvistični izobrazbi sta nemalo prispevala tudi mag. Barry Ridge in dr. Roland Perko – Roli. Barry po angleško, Roli pa po nemško. Obema se zahvaljujem tudi za marsikatero konstruktivno debato na mojo raziskovalno temo, Mateju pa za pomoč pri sistemu Matlab, brez katerega bi si izdelavo eksperimentalnega dela magistrske naloge težko predstavljal.

Da računalništvo predstavlja le neznamenit del celotnega človekovega znanstvenega, tehnološkega in siceršnjega udejstvovanja, mi je najboljše od vseh razodel Jurij Šorli. Zlepa ne bom pozabil pogovorov, v okviru katerih me je izobraževal o najrazličnejših rečeh. Teme pogovorov so segale od nedoumljivih skrivnosti internetne televizije, elektromagnetizma in ameriškega volilnega sistema, pa vse do obskurnosti, kot je npr. problem pretakanja prašnatih snovi po ceveh avtomatov za pripravo kave. Te „seanse“ so mi, upam, vsaj malo razširile obzorje, vsekakor pa sem v njih užival. Hvala, Jurij!

Dušanu Omerčeviču se zahvaljujem za obojestransko koristne matematično obarvane debate. Poleg tega je pokazal, kako je kljub neraziskovalnemu ozadju možno v razmeroma kratkem času doseči znanstveno odličnost, na podlagi česar sem se tudi sam marsičesa naučil.

Seveda se Laboratoriju za umetne vizualne spoznavne sisteme (LUVSS) ne bi

mogel pridružiti, če ne bi prejel vabila. Zahvala za to gre mag. Matjažu Joganu. Zahvaljujem se tudi drugim članom LUVSS, s katerimi sem sodeloval: dr. Ondreju Drbohlavu, mag. Mateju Kristanu in Alenu Vrečku. Vsak od njih je po svoje pripomogel k odličnemu delovnemu vzdušju v laboratoriju.

Za moralno podporo pri izdelavi magistrske naloge se zahvaljujem predvsem staršem in starim staršem. Zaključno poved v zahvali pa naj namenim vsem tistim, ki sem jih (nehote) izpustil, čeprav so k magistrski nalogi morda veliko prispevali.

Povzetek

Magistrska naloga obravnava problem izbire značilnic za samodejno detekcijo objektov dane vizualne kategorije na slikah. Cilj problema detekcije objektov je čim natančneje določiti položaje in velikosti vseh objektov vizualne kategorije na vsaki podani sliki, pri čemer število objektov, upodobljenih na posameznih slikah, ni znano vnaprej. Problem detekcije pogosto otežujejo moteče ozadje na slikah, raznolikost izgleda objektov vizualne kategorije, raznolikost velikosti upodobitev objektov in delna zakrivanja.

V magistrski nalogi se osredotočamo na sisteme, ki pred izvedbo detekcijskega postopka reprezentirajo (predstavijo) vsako vhodno sliko z neko množico značilnic, pripravljeno v dveh stopnjah učne faze: prva stopnja pridobi na osnovi neke učne slikovne zbirke izhodiščno množico značilnic, v drugi stopnji pa se iz pridobljene množice izberejo značilnice za uporabo pri detektiranju objektov na testnih slikah. Magistrska naloga se prvenstveno posveča drugi stopnji, torej izbiri značilnic. Primerna izbirna metoda zmanjša računsko zahtevnost detektiranja objektov, poleg tega pa odstrani značilnice, ki so za detekcijo objektov dane kategorije nekoristne ali celo škodljive.

V magistrski nalogi predstavljamo dve izbirni metodi. Jedro prve, t.i. filtrske metode, predstavlja funkcija za ocenjevanje značilnic, ki temelji na prevedbi problema detekcije objektov na učnih slikah v problem dvorazredne klasifikacije učnih slik. Ker se večina obstoječe teorije in prakse izbire značilnic nanaša na klasifikacijsko domeno, tovrstna problemska prevedba bistveno razširi razpon izbirnih pristopov, ki se jih je mogoče poslužiti. Funkcija za ocenjevanje značilnic je zasnovana tako, da jo je mogoče enostavno vgraditi v izbirno ogrodje AdaBoost, ki izbira značilnice z implicitnim upoštevanjem njihove medsebojne odvisnosti.

Druga predstavljena izbirna metoda temelji na konceptu notranje optimizacije. Metoda sistematično poganja detekcijski postopek na učnih slikah za različne kandidatne množice značilnic. Cilj takega pristopa je poiskati množico značilnic, ki vodi do najvišje detekcijske točnosti na učni slikovni množici. Tako zastavljen problem je možno enostavno preoblikovati v problem iskanja optimalnega vozlišča v grafu prostora stanj.

V izvedbi sistema smo vsako od obeh izbirnih metod povezali s preizkušenim detekcijskim pristopom in z dvema sodobnima in obetavnima metodama za pridobivanje značilnic. Preizkusili smo vse štiri kombinacije metod za pridobivanje značilnic in izbirnih metod. Za vrednotenje rezultatov preizkusov smo uporabili pet testnih slikovnih zbirk, od katerih vsaka opredeljuje neko vizualno kategorijo in samostojno detekcijsko nalogo. Izkazalo se je, da v nekaterih primerih že zelo majhno število izbranih značilnic zadošča za dosego precej visoke detekcijske točnosti. Pri kombinaciji filtrske izbirne metode, metode za pridobivanje značilnic avtorjev Fidler in sod. (2006) in testne zbirke avtomobilov UIUC že štiri značilnice, izbrane izmed 952 kandidatov, vodijo do detekcijske točnosti, primerljive z najvišjo do sedaj objavljeno točnostjo za zbirko UIUC.

Ključne besede: računalniški vid, izbira značilnic, detekcija objektov, postopek AdaBoost, pridobivanje značilnic.

Abstract

This master's thesis deals with the problem of selecting features for automatic detection of objects of a given visual category in images. The goal of object detection is to determine the locations and sizes of all objects of the given category in each input image, where the number of objects displayed in individual images is not known in advance. Various circumstances may contribute to the difficulty of object detection, such as cluttered backgrounds, diversity in the appearance of objects in the category, diversity in the scale of the displayed objects, and partial occlusions.

The thesis primarily focuses on systems that represent each input image by a set of features and use the resulting image representations to form detection hypotheses. The feature set is assumed to be prepared by a two-step procedure in the training stage. In the first step, the initial feature set is extracted from a set of training images. In the second step, the features to be used in the test stage are selected from the initial set. The thesis is devoted to the second step, i.e., feature selection. An effective selection method reduces the computational complexity in the test stage and eliminates features that are useless or even harmful for detecting objects of a given category.

In the master's thesis, we present two feature selection methods. At the core of the first one (the so-called filter method) is a feature evaluation function that is based on a transformation of the problem of detecting objects in training images into a problem of classifying training images themselves. Since much existing theory and practice of feature selection pertains to the classification rather than detection domain, such a problem transformation greatly expands the range of applicable selection approaches. Because of its design, the feature evaluation function can be straightforwardly incorporated into the AdaBoost selection framework, which selects features by considering their interdependence in an implicit manner.

The second selection method that is presented in this thesis (the so-called wrapper method) systematically runs the detection procedure on a fixed training image set for various candidate feature sets. The goal of such a selection approach is to find a feature set that enables the detection procedure to attain the highest detection accuracy on the training image set. The problem of finding the optimal candidate feature set can be easily transformed into a problem of finding the optimal vertex in the corresponding state space graph.

In the implementation of the system, each of the two selection methods was integrated with a well-known detection approach and with two promising state-of-the-art feature extraction methods. We experimentally tested all four combinations of feature extraction and feature selection methods. The experiments were evaluated using five test image sets, each of which defined a visual category and an independent detection setup. In some cases, as it turned out, a very small number of selected features suffices to bring about a fairly high detection accuracy. Particularly remarkable results were achieved for the combination of the filter selection method, the feature extraction method due to Fidler et al. (2006), and the UIUC

car dataset. In this setup, the selection of merely four features out of 952 initial candidates led to a detection accuracy that is comparable to the state-of-the-art results for the UIUC dataset.

Key words: computer vision, feature selection, object detection, AdaBoost algorithm, feature extraction.

Kazalo

1	Uvod	1
1.1	Opredelitev ključnih pojmov in motivacija	1
1.1.1	Detekcija objektov vizualnih kategorij	1
1.1.2	Lokalne značilnice	3
1.1.3	Motivacija koncepta izbire značilnic	4
1.2	Izbira značilnic v klasifikacijskih in detekcijskih shemah	7
1.2.1	Problem klasifikacije	7
1.2.2	Izbira značilnic v klasifikacijskih shemah	9
1.2.3	Izbira značilnic v detekcijskih shemah	15
1.3	Magistrska naloga	20
1.3.1	Pregled	20
1.3.2	Pričakovani prispevek k znanosti	21
1.3.3	Implementacijske opombe	21
1.4	Zgradba nadaljnjega besedila	22
2	Izbira značilnic	23
2.1	Uvod	23
2.2	Pričakovani vhod in izhod detekcijskega modula	24
2.3	Filtrska metoda za izbiro značilnic	24
2.3.1	Prevedba problema detekcije v problem klasifikacije	25
2.3.2	Funkcija za ocenjevanje značilnic	28
2.3.3	Vgradnja ocenjevalne funkcije v izbirni postopek AdaBoost	30
2.4	Notranjeoptimizacijska metoda za izbiro značilnic	33
2.4.1	Uvod	33
2.4.2	Prostor stanj	33
2.4.3	Iskalne strategije	34
2.4.4	Smer preiskovanja grafa stanj	37
2.4.5	Končne opombe	37
2.5	Povzetek poglavja	38

3	Detekcijski modul	39
3.1	Uvod	39
3.2	Detekcija objektov konstantne velikosti	40
3.2.1	Učenje detekcijskega modela	40
3.2.2	Osnovni postopek detekcije	42
3.2.3	Dušenje sekundarnih hipotez	44
3.3	Detekcija objektov poljubne velikosti	47
3.3.1	Splošna shema večvelikostne detekcije	48
3.3.2	Večvelikostna detekcija pri velikostno invariantnih značilnicah	50
3.4	Izhod detekcijskega postopka	54
3.5	Vrednotenje rezultatov detekcije	55
3.5.1	Ugotavljanje pravilnosti detekcijskih hipotez	55
3.5.2	Ocenjevanje detekcijske točnosti	57
3.6	Povzetek poglavja in končne opombe	59
3.6.1	Povzetek poglavja	59
3.6.2	Končne opombe	60
4	Značilnice	61
4.1	Uvod	61
4.2	Značilnice Fidler in sod.	62
4.2.1	Uvod	62
4.2.2	Predstavitev značilnic	63
4.2.3	Reprezentacija slik s podanimi značilnicami	64
4.2.4	Učenje hierarhije značilnic	69
4.2.5	Uporaba značilnic Fidler in sod. v drugih modulih sistema	72
4.3	Značilnice Ferrari in sod.	72
4.3.1	Uvod	72
4.3.2	Skupine k robnih odsekov	73
4.3.3	Učenje značilnic	76
4.3.4	Reprezentacija slik s podanimi značilnicami	78
4.3.5	Uporaba značilnic Ferrari in sod. v drugih modulih sistema	78
4.4	Povzetek poglavja	78
5	Eksperimentalni rezultati	81
5.1	Testne slikovne množice	81
5.2	Podrobnosti izvajanja preizkusov	83
5.2.1	Shema izvajanja preizkusov glavnega niza	83
5.2.2	Filtrska izbira značilnic Fidler in sod.	86
5.2.3	Filtrska izbira značilnic Ferrari in sod.	90
5.2.4	Notranjeoptimizacijska izbira	92
5.2.5	Vrednosti nekaterih parametrov	94
5.3	Rezultati filtrske izbire značilnic Fidler in sod.	94
5.3.1	Rezultati preizkusov	94
5.3.2	Delovanje postopka AdaBoost	102

5.4	Rezultati notranjeoptimizacijske izbire značilnic Fidler in sod. . . .	107
5.5	Rezultati filtrske izbire značilnic Ferrari in sod.	111
5.6	Rezultati notranjeoptimizacijske izbire značilnic Ferrari in sod. . .	115
5.7	Povzetek poglavja	118
6	Zaključek	123
6.1	Pregledni povzetek magistrske naloge	123
6.2	Smeri nadaljnjega razvoja	125

Slike

1.1	Ilustracija velikostne in rotacijske invariantnosti značilnic	5
1.2	Pregledna shema izbire značilnic za detekcijo	6
1.3	Vrednotenje značilnic pri metodi Viola in Jones (2004)	16
2.1	Shema filtrske izbirne metode	25
2.2	Porazdelitev $S(Z, I)$ za izbrano značilnico Z na učnih slikah I . . .	29
2.3	Shema notranjeoptimizacijske izbirne metode	34
2.4	Primer grafa prostora stanj	35
3.1	Ilustracija osnovnega učnega in testnega postopka pri enovelikostni detekciji	41
3.2	Primer prostora detekcijskih hipotez po izvedbi enovelikostnega glasovanja	44
3.3	Motivacija za uvedbo algoritma MDL	47
3.4	Ilustracija učenja in testiranja večvelikostnega detekcijskega modela	51
3.5	Preverjanje pravilnosti detekcijskih hipotez	56
3.6	Primer krivulje RPC	58
4.1	Edina značilnica Fidler in sod. na prvi ravni	64
4.2	Preverjanje nastopa značilnice Fidler in sod. v sliki	65
4.3	Primer skupine k robnih odsekov	75
5.1	Primeri slik iz vseh uporabljenih testnih zbirk	85
5.2	Primeri učnih slik iz \mathcal{I}_{zd} (filtrska izbirna metoda, značilnice Fidler in sod.)	87
5.3	Po dva primera pozitivnih in negativnih slik učne zbirke UIUC	88
5.4	Primeri slik iz zbirk GRAZ-01 in LPP-34 ter iz raznih spletnih virov	89
5.5	Primeri avtomobilov iz Yahoojeve slikovne zbirke	92
5.6	Primeri „pozitivnih“ učnih slik za notranjeoptimizacijsko izbiro	93
5.7	Odvisnost F_{\max} od števila izbranih značilnic (filtrska izbira, Fidler in sod.)	96

5.8	Krivulje RPC za prvih 5, 10 in 50 izbranih značilnic (filtrska izbira, Fidler in sod.)	98
5.9	Detekcijski model in manifestacije posameznih prvoizbranih značilnic (filtrska izbira, Fidler in sod.)	101
5.10	Primeri pravih, napačnih in manjkajočih detekcij pri optimalnem pragu τ (filtrska izbira, Fidler in sod.)	103
5.11	Primer zglajenega hipotetnega prostora pri detekciji avtomobilov	104
5.12	Stopnje napake izbranih statičnih in dinamičnih optimalnih klasifikatorjev (filtrska izbira, Fidler in sod.)	106
5.13	Klasifikacijska točnost prvih treh izbranih optimalnih klasifikatorjev (filtrska izbira, Fidler in sod.)	108
5.14	Spreminjanje uteži učnih slik v prvih štirih iteracijah postopka AdaBoost (filtrska izbira, Fidler in sod.)	109
5.15	Odkvisnost F_{\max} od števila izbranih značilnic (notranja optimizacija, Fidler in sod.)	110
5.16	Odkvisnost ocene izbrane množice značilnic od števila izbranih značilnic (notranja optimizacija, Fidler in sod.)	111
5.17	Odkvisnost F_{\max} od števila izbranih značilnic (filtrska izbira, Ferrari in sod.)	114
5.18	Primeri manifestacij značilnic Ferrari in sod. na testnih slikah	117
5.19	Odkvisnost F_{\max} od števila izbranih značilnic (notranja optimizacija, Ferrari in sod.)	119
5.20	Odkvisnost ocene izbrane množice značilnic od števila izbranih značilnic (notranja optimizacija, Ferrari in sod.)	120

Tabele

4.1	Primeri upodobitev značilnic Fidler in sod. in njihovih komponent	66
4.2	Primeri značilnic Ferrari in sod. in pripadajočih gruč skupin k robnih odsekov	77
5.1	Izbrani podatki o testnih slikovnih množicah	83
5.2	Izbrani podatki o učnih množicah \mathcal{I}_{zd} (filtrska izbirna metoda, značilnice Fidler in sod.)	88
5.3	Število pozitivnih in negativnih slik v \mathcal{I}_i (filtrska izbira, Fidler in sod.)	89
5.4	Razpon velikostne piramide v testni fazi (filtrska izbira, Fidler in sod.)	91
5.5	Nekateri podatki o množicah \mathcal{I}_{zd} (filtrska izbira, Ferrari in sod.) . .	91
5.6	Vrednosti izbranih parametrov metode.	95
5.7	Detekcijski rezultati za zbirko UIUC, kot jih navajajo nekateri drugi avtorji	97
5.8	Prvih 10 izbranih značilnic (filtrska izbira, Fidler in sod.)	99
5.9	Vpliv algoritma MDL na detekcijsko točnost	104
5.10	Prvih 10 značilnic Fidler in sod., izbranih po padajočih statičnih ocenah	105
5.11	Primerjava med dinamičnimi in statičnimi ocenami značilnic Fidler in sod.	106
5.12	Prvih 10 izbranih značilnic (notranja optimizacija, Fidler in sod.) .	112
5.13	Prvih 10 izbranih značilnic (filtrska izbira, Ferrari in sod.)	115
5.14	Prvih 10 značilnic Ferrari in sod., izbranih po padajočih statičnih ocenah	116
5.15	Vpliv večvelikostne obravnave testnih slik pri uporabi značilnic Ferrari in sod. na detekcijsko točnost	116
5.16	Prvih 10 izbranih značilnic (notranja optimizacija, Ferrari in sod.)	118
5.17	Pregled detekcijskih točnosti F_{\max} , dobljenih za posamezne kombinacije izbirnih metod, sestavov značilnic, testnih zbirk in moči izbirnih množic	120

5.18 Pregled detekcijskih točnosti EER, dobljenih za posamezne kombinacije izbirnih metod, sestavov značilnic, testnih zbirk in moči izbirnih množic	121
--	-----

Algoritmi

2.1	Izbirni postopek AdaBoost	32
2.2	Izbira značilnic s požrešnim preiskovanjem prostora stanj	36
2.3	Izbira značilnic s snopovnim preiskovanjem prostora stanj	36
3.1	Algoritem MDL pri enovelikostni detekciji	46
4.1	Preverjanje nastopa značilnice Fidler in sod. v sliki	68

Uvod

V uvodnem poglavju najprej, v razdelku 1.1, na kratko predstavimo in motiviramo obe naslovni tematiki, detekcijo objektov dane vizualne kategorije na slikah in izbiro značilnic za detekcijo objektov. Sledi razdelek 1.2, v katerem prikažemo problem izbire značilnic v širšem kontekstu in predstavimo nekatera sorodna dela. Razdelek 1.3 namenjamo kratkemu pregledu magistrske naloge in izpostavitvi pričakovanega prispevka naloge. Poglavje zaključujemo z razdelkom 1.4, ki služi kot kazalo po nadaljnjih poglavjih magistrske naloge.

1.1 Opredelitev ključnih pojmov in motivacija

„Preden uporabiš tujko, si zamisli, da si jo kupil za devize,“ svetuje Jože Šircelj v duhovitem in poučnem *Žepnem bontonu* iz leta 1964, ko so devize povprečnemu Slovincu predstavljale nadvse želeno, a težko dostopno dobrino. Zdi se, kot da smo količino jezikovnih „deviz“, ki bi si jo po *Žepnem bontonu* nemara še smeli privoščiti, potrošili že s samim naslovom magistrske naloge. Zakaj govorimo o „izbiri značilnic za detekcijo objektov dane vizualne kategorije“ in ne morda o „izbiri značilnic za zaznavanje predmetov danega razreda, opredeljenega glede na skupne vidne lastnosti“? Zaradi želje po čim natančnejšem izražanju in zaradi sledenja smernicam, ki se polagoma uveljavljajo pri prevajanju angleškega znanstvenega izrazoslovja v slovenščino. Za porabljene „devize“ se bomo skušali odkupiti z razlago vseh navedenih tujk.

1.1.1 Detekcija objektov vizualnih kategorij

Detekcija objektov na slikah (angl. object detection) je področje znotraj discipline *računalniškega vida* oziroma, širše, *razpoznavanja vzorcev*. Ukvarja se z iskanjem vzorcev z želenimi, pogosto implicitno podanimi lastnostmi v slikah. Primerke vzorcev z želenimi lastnostmi, ki nastopajo v dani sliki, imenujemo *objekti*; vsi preostali vzorci v sliki tvorijo nezanimivo *ozadje*. V okviru detekcije objektov je pojem objekta mnogo širši od pojma predmeta, kot ga običajno razumemo v slovenskem jeziku. Objekt je definiran preprosto kot vzorec, čigar primerke iščemo na dani sliki. Lahko gre res za predmet (npr. avtomobil, televizor, knjiga), vendar

lahko za „objekt“ proglasimo praktično karkoli. Zanima nas lahko iskanje podob človeških obrazov, pešcev, določenih živalskih vrst, geometrijskih vzorcev itn.

Pri detekciji objektov se običajno omejimo na iskanje objektov z določenimi skupnimi *vizualnimi* lastnostmi. (Med raziskovalci na področju razpoznavanja vzorcev obstaja neformalen konsenz, da je beseda „vizualen“ točnejši prevod angleškega „visual“ kot beseda „viden“.) Takšni objekti pripadajo neki *vizualni kategoriji*. Primeri vizualnih kategorij so osebni avtomobil, konj, skodelica, velika tiskana črka A, človeški obraz itn. Primerki osebnih avtomobilov se med seboj precej razlikujejo, vendar imajo vsi določene skupne vizualne lastnosti (kolesa, luči, streha ipd.). Medtem ko bi kategorijo „zgradba“ verjetno še lahko označili kot vizualno kategorijo, je kategorija „prevozno sredstvo“ preveč raznolika, da bi ji mogli pripisati to oznako. V nadaljevanju bomo besedno zvezo „objekt dane vizualne kategorije“ običajno skrčili kar v besedo „objekt“, saj bomo obravnavali le detekcijo objektov neke vnaprej določene (čeprav poljubne) vizualne kategorije.

Pojem *detekcije* objektov ima v sferi razpoznavanja vzorcev natančno določen pomen, ki ga ne gre zamenjevati z (ohlapnejšimi) pojmi, kot so „zaznavanje“ in „prepoznavanje“ objektov na slikah. Problem detekcije objektov dane vizualne kategorije na podani sliki je formuliran na sledeči način:

1. Cilj je določiti položaje in velikosti vseh objektov dane kategorije na dani sliki.
2. Slika lahko vsebuje poljubno število objektov, kar vključuje tudi možnost, da slika ne vsebuje nobenega objekta.
3. Število objektov na sliki ni znano vnaprej.

Pri točki 1 včasih zahtevamo tudi določitev orientacij objektov, včasih pa nas niti velikosti ne zanimajo in se osredotočimo le na položaje. V magistrski nalogi se bomo ukvarjali z določanjem položajev in velikosti objektov. Točka 3 je pomembna postavka, ki ločuje problem detekcije od enostavnejšega problema *lokalizacije*. Detekcijski postopek se ne more opreti na nikakršno predpostavko glede števila objektov na sliki, saj vnaprej ni znano niti to, ali slika sploh vsebuje kakšen objekt dane kategorije.

Detekcijo ponavadi otežkočajo raznovrstne okoliščine:

- Vizualna kategorija lahko izkazuje precejšnjo raznolikost izgleda objektov. Avtomobili, na primer, se razlikujejo po znamkah in karoserijskih oblikah. Velika tiskana črka A lahko nastopa v različnih pisavah. Pri fleksibilnih objektih je raznolikost še večja: konj je lahko upodobljen v mirujočem stanju, v galopu, na zadnjih nogah, s sklonjeno glavo, z zelo različnimi legami nog itn.
- Objekti so na slikah lahko upodobljeni v zelo različnih velikostih.
- Objekti so lahko upodobljeni v različnih zavrtenostih okrog osi, pravokotne na slikovno ravnino.

- Objekti so lahko deloma zakriti.
- Objekti se lahko deloma prekrivajo.
- Slike lahko vsebujejo poljubno količino motečega ozadja ali „šare“ (angl. clutter).
- Objekti so na slikah lahko prikazani v različnih pogledih, npr. v tlorisu, narisu ali stranskem risu. Ti pogledi se običajno med seboj bistveno razlikujejo, zato jih je možno obravnavati kot različne vizualne kategorije. Leibe (2004), na primer, obravnava detekcijo avtomobilov v pogledu od strani in detekcijo avtomobilov v pogledu od zadaj kot dva ločena detekcijska problema.

Zaradi naštetih dejavnikov je splošna detekcija zahteven problem, ki še zdaleč ni zadovoljivo rešen.

Detekcijski postopek pri detektiranju objektov dane kategorije uporablja *detekcijski model* za to kategorijo. V nekaterih omejenih domenah je detekcijski model vgrajen v samo detekcijsko metodo. Tak model je določen že v fazi gradnje sistema. V magistrski nalogi nas bodo pa bolj zanimali modeli, ki so *naučeni* na podlagi učnih podatkov. V takšnih primerih je detekcijska metoda razdeljena na dve ločeni fazi:

Učna faza, v kateri se na podlagi množice *učnih slik* za dano vizualno kategorijo zgradi detekcijski model kategorije.

Testna faza, v kateri se naučeni detekcijski model preizkusi nad množico *testnih slik*, ki mora biti povsem ločena od učne množice. Z drugimi besedami: nobena učna slika ne sme nastopati v testni fazi.

Zahteva o ločitvi učne in testne množice je ena od najpomembnejših postavk, ki zadevajo učne algoritme. Sicer učna množica ni omejena z ničimer. Vsebovati sme poljubno količino slik s poljubnimi lastnostmi. Kot testno množico pa pogosto obravnavamo katero od pripravljenih slikovnih zbirk, da lahko dobljene rezultate primerjamo z rezultati drugih metod.

Obstajajo tudi drugačne paradigme učenja, s katerimi pa se v magistrski nalogi ne bomo ukvarjali. Omenimo paradigmo *nenehnega učenja* (angl. continuous learning), pri kateri delitev na učno in testno fazo ni smiselna, saj se učni model nenehno nadgrajuje in obnavlja (Skočaj in sod., 2007).

1.1.2 Lokalne značilnice

Večina detekcijskih metod pri gradnji detekcijskega modela in pri njegovem preizkušanju uporablja množico *lokalnih značilnic* (angl. local features), s katerimi *reprezentira* učne in testne slike. Lokalne značilnice so občutljive na določene lokalne lastnosti ali strukture v sliki. Gaborjev filter, prikazan na sliki 4.1 na strani 64, je primer značilnice, občutljive na lokalno spremembo intenzitete v vodoravni smeri.

Z izrazom *reprezentacija slike* označujemo postopek izražave slike z dano množico značilnic, pa tudi rezultat takšne izražave, torej predstavitev slike z množico značilnic. Reprezentacija posameznim lokalnim strukturam v sliki priredi ustrezajoče značilnice. Reprezentacija je lahko *enolična* ali *neenolična*. Pri enolični reprezentaciji se vsaki lokalni strukturi lahko priredi kvečjemu ena značilnica. Pri neenolični reprezentaciji se lahko več značilnic priredi isti lokalni strukturi. Neenolična reprezentacija lokalne strukture je pogosto opremljena z vektorjem, ki podaja uteži ali verjetnosti prirejanj posameznih značilnic tej lokalni strukturi.

Če značilnica pri reprezentaciji določene lokalne strukture sodeluje z neničelno utežjo izražave, pravimo, da se značilnica v sliki *manifestira* na položaju te strukture oziroma na slikovnih koordinatah, na katerih se nahaja ta struktura. Na položaju dane lokalne strukture se torej manifestirajo vse značilnice, ki jih reprezentacijski postopek priredi tej strukturi. Manifestacijo značilnice v sliki bomo označevali tudi z enakovrednima izrazoma „pojavitev“ in „nastop“. Seveda se lahko ista značilnica v sliki manifestira na več različnih mestih.

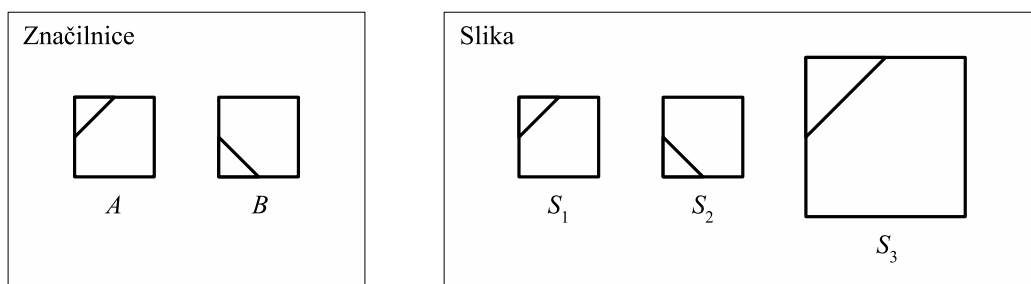
Reprezentacija slike s podano množico značilnic ni nič drugega kot seznam manifestacij značilnic iz podane množice na sliki. Vsaka manifestacija je običajno opremljena vsaj s podatki o identiteti in položaju manifestirane značilnice. V primeru neenolične reprezentacije je manifestaciji smiselno pripisati tudi utež, s katero značilnica nastopa pri izražavi lokalne strukture. V nekaterih primerih pa se manifestaciji lahko priredi tudi velikost ali orientacija, kot je razloženo v sledečem odstavku.

Naj besedna zveza *sestav značilnic* označuje shemo zgradbe značilnic, ki določa njihove skupne lastnosti. Sestav značilnic je *velikostno invarianten*, če je manifestacijam značilnic smiselno prirediti (relativne) velikosti, ki se spreminjajo sorazmerno s spreminjanjem velikosti lokalnih struktur, v katerih se značilnice manifestirajo. Če se velikostno invariantna značilnica Z na sliki manifestira na položaju (x, y) , v velikosti t in z utežjo w , se bo po k -kratnem povečanju slike značilnica Z manifestirala na položaju (kx, ky) , v velikosti kt in z nespremenjeno utežjo w . Na podoben način je možno definirati *rotacijsko invariantnost*. Manifestacijam rotacijsko invariantnih značilnic je smiselno prirediti (relativne) orientacije, ki se spreminjajo skladno z zavrtenostjo lokalnih struktur, v katerih se značilnice manifestirajo.

Pojem velikosti je smiselno definirani le za velikostno invariantne značilnice in njihove manifestacije. Značilnice, ki te lastnosti nimajo, so običajno zasnovane tako, da se ponovljivo odzivajo le na lokalne strukture določenih velikosti. Podobno velja za pojem orientacije in rotacijsko invariantne značilnice. Predstavljeni koncepti so ilustrirani na sliki 1.1.

1.1.3 Motivacija koncepta izbire značilnic

V večini sodobnih metod za detekcijo objektov je množica značilnic, namenjena za reprezentiranje učnih in testnih detekcijskih slik, dobljena z uporabo nekega postopka za *pridobivanje značilnic* (angl. feature extraction) na podlagi podanih



Slika 1.1: Množica značilnic (levo) in lokalne strukture na sliki (desno). Če sestav značilnic ni niti velikostno niti rotacijsko invarianten, je lokalna struktura S_1 manifestacija značilnice A , S_2 je manifestacija B , na položaju strukture S_3 pa ne nastopa nobena od značilnic A in B . Če je sestav značilnic velikostno, ne pa tudi rotacijsko invarianten, je S_3 manifestacija značilnice A v relativni velikosti 2, če se kot velikostna mera značilnice obravnava npr. dolžina stranice. Če je sestav značilnic rotacijsko invarianten, je lokalna struktura S_1 manifestacija značilnice A v orientaciji 0° in hkrati manifestacija B v orientaciji 270° . Podobno velja za S_2 . V primeru rotacijske invariantnosti je ena od obeh značilnic dejansko odveč, saj se značilnici razlikujeta le po zavrtenosti.

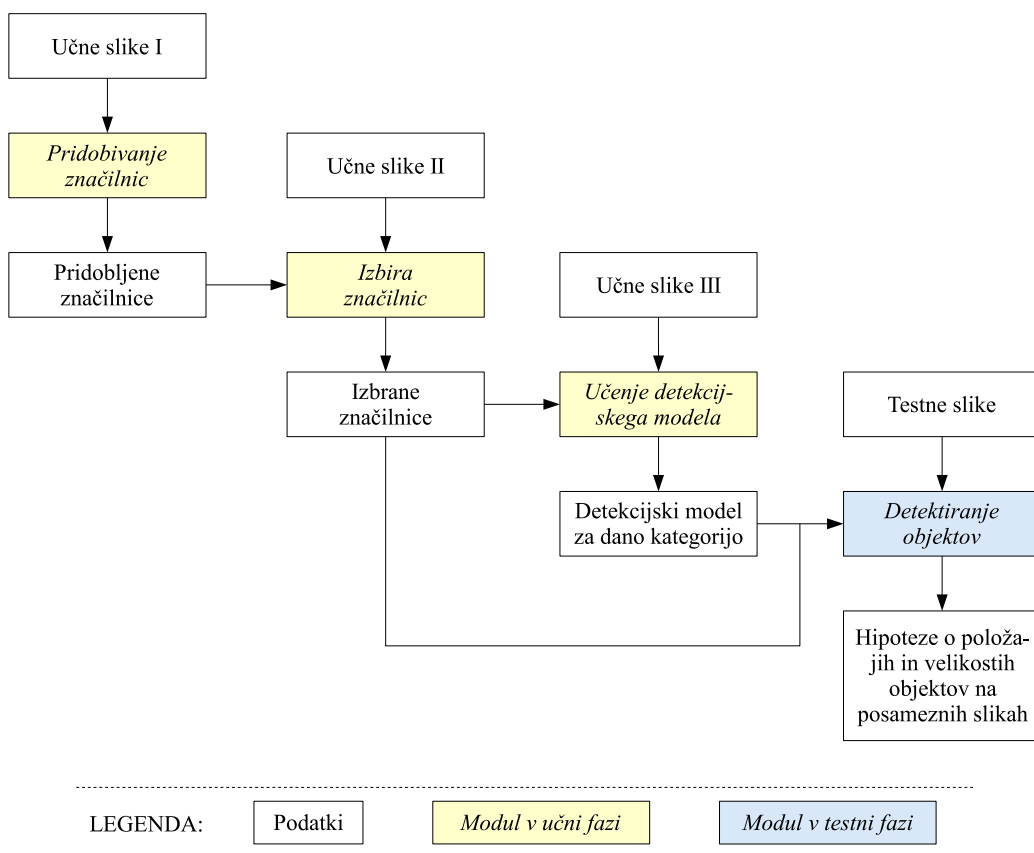
slik. Vendar je množica tako nastalih značilnic pogosto precej velika, poleg tega pa ni nujno, da so vse značilnice dejansko koristne za detekcijo objektov dane kategorije. Še več, upoštevanje nekaterih značilnic utegne celo zmanjšati detekcijsko sposobnost metode; takšne značilnice so torej ne le nekoristne, marveč celo škodljive za detekcijo.

Množica značilnic za reprezentacijo učnih in testnih slik za detekcijo se zato običajno pripravi v dveh stopnjah. Prva stopnja obsega pridobivanje značilnic na podlagi neke slikovne množice. V drugi stopnji se iz množice pridobljenih značilnic *izberejo* značilnice za uporabo v sami detekcijski metodi. Postopek za izbiro značilnic (angl. feature selection) se pri svojem izvajanju prav tako opira na neko množico slik: množico za pridobivanje značilnic, množico za izbiranje značilnic, množico za učenje detekcijskega modela in testno množico, ki služi preizkušanju naučenega detekcijskega modela. Učne množice se med seboj lahko v poljubni meri prekrivajo. Spoštovati je treba le pogoj, da je testna množica povsem ločena od vseh treh učnih množic. Sosledje pridobivanja značilnic, izbire značilnic, učenja detekcijskega modela in uporabe naučenega modela je shematsko prikazano na sliki 1.2. Izraz *modul*, ki ga izpostavljam na sliki, naj predstavlja kakršnokoli izvajalno enoto. Modul lahko sestoji iz več manjših modulov.

Vloga izbire značilnic je torej večplastna. Zmanjšanje količine značilnic vodi do večje časovne in prostorske učinkovitosti pri izvajanju detekcijskega postopka. Poleg tega utegne primeren izbirni postopek izločiti veliko značilnic, ki so za detekcijo nekoristne ali celo škodljive, kar pripomore k izboljšanju detekcijske sposobnosti

oziroma k povečanju detekcijske točnosti v testni fazi.

Viola in Jones (2004) sta pokazala, da ustrezen izbirni postopek omogoča uporabo razmeroma preproste metode za pridobivanje značilnic. Metoda za detekcijo obrazov, ki sta jo zasnovala, po preprostih pravilih pridobi veliko količino enostavnih značilnic, ki so povečini videti povsem nekoristne za detekcijo obrazov. Vendar pa sta Viola in Jones z uporabo izbirne strategije, temelječe na postopku AdaBoost (Freund in Schapire, 1997), odkrila, da lahko že zelo majhno število ustrezno izbranih značilnic vodi do presenetljivo visoke detekcijske točnosti in hkrati omogoča detekcijo obrazov v realnem času. Metodo Viole in Jonesa, ki



Slika 1.2: Pregledna shema izbire značilnic za detekcijo. Učna faza zajema pridobivanje značilnic, izbiro značilnic in učenje detekcijskega modela. Izbrane značilnice se uporabljajo pri učenju detekcijskega modela in v testni fazi. Naj opozorimo, da je prikazana shema v prvi vrsti namenjena ilustraciji pojmov in predstavlja določeno poenostavitev obeh izbirnih metod, predstavljenih v poglavju 2. V obeh metodah modul za izbiro značilnic uporablja naučeni detekcijski model za posamezne značilnice in se poslužuje funkcij modula za detektiranje objektov. Vendar je te odvisnosti možno skriti v sam modul „Izbira značilnic“ in tedaj shema konceptualno ustreza predstavljenima metodama.

predstavlja pomemben navdih za magistrsko nalogo, bomo podrobneje predstavili v odseku 1.2.3.1.

1.2 Izbira značilnic v klasifikacijskih in detekcijskih shemah

V tem razdelku bomo problem izbire značilnic postavili v širši okvir. V ta namen bomo iz domene detekcije objektov začasno prešli na domeno *klasifikacije primerov v razrede*, saj se večina teorije izbire značilnic nanaša na klasifikacijske probleme. V podrazdelku 1.2.1 bomo predstavili problem klasifikacije. V podrazdelku 1.2.2 se bomo posvetili pregledu nekaterih pristopov k izbiri značilnic za klasifikacijo. V podrazdelku 1.2.3 se bomo vrnili na področje detekcije in predstavili izbiro značilnic pri treh različnih detekcijskih metodah.

1.2.1 Problem klasifikacije

1.2.1.1 Splošen problem klasifikacije

Problem klasifikacije primerov v razrede je ena od osrednjih tematik strojnega učenja (Mitchell, 1997; Kononenko, 2005), pomembno mesto pa zaseda tudi v domeni razpoznavanja vzorcev (Duda in sod., 2000).

Cilj splošnega problema klasifikacije je razvrščanje *primerov v razrede*. Pri medicinski diagnostiki bi posamezni pacienti ustrezali primerom, diagnoza pa razredu. Primeri so opisani z *atributi* in njihovimi *vrednostmi*. Atributi so običajno diskretne ali zvezne spremenljivke. Primer diskretnega atributa je barva las (dovoljene vrednosti so npr. črna, rjava, plava, rdeča). Primer zveznega atributa je temperatura vode (množica dovoljenih vrednosti je npr. zvezni interval $[0, 100]$).

Razred ni nič drugega kot vrednost *ciljnega atributa*, ki je v domeni klasifikacije vselej diskretna spremenljivka. Pri medicinski diagnostiki bi ciljnemu atributu oz. razredu ustrezala diagnoza. Če je ciljni atribut dvovrednostna spremenljivka, gre za dvorazredni ali tudi *dvojiški* klasifikacijski problem. V tem primeru je pogosto eden od obeh razredov opredeljen kot *pozitivni razred*, drugi pa kot *negativni razred*. Primeri, ki pripadajo pozitivnemu oz. negativnemu razredu, so označeni kot pozitivni oz. negativni primeri.

Celotna množica primerov, opisanih z vrednostmi atributov, je običajno razdeljena na učno in testno množico. Pri učnih primerih so vrednosti vseh atributov, vključno s ciljnim atributom, znane. (V realnih problemih lahko sicer nekatere vrednosti atributov za posamične primere manjkajo.) Učni primeri so namenjeni učenju (indukciji) *klasifikatorja*, torej učenju preslikave iz prostora vrednosti atributov v množico razredov oz. vrednosti ciljnega atributa. Naučeni klasifikator se v testni fazi preizkusi na množici testnih primerov. Klasifikator mora za vsak testni primer na podlagi vrednosti atributov napovedati razred primera, napovedani razredi pa se potem primerjajo z dejanskimi razredi primerov. Na podlagi

rezultatov primerjav se izračuna *točnost* klasifikatorja, ki predstavlja enostavno mero klasifikacijske uspešnosti.

V strojnem učenju se izraz „značilnica“ običajno nanaša na kombinacijo atributa in njegove vrednosti za določen primer. „Barva“ je torej atribut, „barva = modra“ pa bi lahko bila značilnica nekega primera. Vendar pa se na področju izbire značilnic pojem „značilnica“ praviloma uporablja kot sopomenka za izraz „atribut“. Tej terminologiji bomo sledili tudi mi. Zaradi konsistentnosti pa bomo v nadaljevanju uporabljali izključno izraz *značilnica*. „Barva“ naj bo torej značilnica, formulacija „barva = modra“ pa naj ponazarja vrednost značilnice „barva“ pri danem primeru.

1.2.1.2 Vizualna klasifikacija in lokalne značilnice

Zdi se, kot da je koncept značilnice v smislu klasifikacije primerov povsem nepovezan s konceptom lokalnih značilnic iz podrazdelka 1.1.2. Vendar pa sta koncepta enakovredna. To trditev bomo skušali utemeljiti na primeru vizualne klasifikacije slik z uporabo lokalnih značilnic.

Problem vizualne klasifikacije je ponavadi definiran na sledeči način. Dana je množica vizualnih kategorij. Vsaka slika prikazuje objekt, ki pripada eni od kategorij. To velja tako za učne kot tudi za testne slike. Naloga učnega algoritma je zgraditi klasifikator na podlagi učnih slik, reprezentiranih z lokalnimi značilnicami, in oznak kategorij, ki jim pripadajo posamezne slike. Naučeni klasifikator se nato preizkusi na množici testnih slik; njegova naloga je uvrstiti vsako sliko v ustrezno kategorijo. Klasifikatorjeve napovedi se primerjajo z dejanskimi podatki o kategorijski pripadnosti posameznih testnih slik.

Precej enostavno je vzpostaviti povezavo med tako definirano vizualno klasifikacijo in splošno klasifikacijo: slikam ustrezajo primeri; vizualnim kategorijam ustrezajo razredi; lokalnim značilnicam ustrezajo značilnice v splošni klasifikacijski domeni. Edino težavo predstavljajo *vrednosti* značilnic za posamezne primere (slike). Pri medicinski diagnostiki, npr., bi bila vrednost zvezne značilnice „temperatura“ za nekega pacienta (= klasifikacijski primer) lahko enaka 38,1 °C, vrednost dvojiške značilnice „prebolel bolezen X“ bi znašala bodisi 1 bodisi 0 itn. Učni algoritem gradi klasifikator prav na podlagi vrednosti značilnic za posamezne primere. Čemu bi ustrezale vrednosti značilnic pri vizualni klasifikaciji?

Najpreprostejša definicija vrednosti lokalne značilnice Z bi bila takšna: Če značilnica Z nastopa v sliki, je njena vrednost za to sliko (= klasifikacijski primer) enaka 1, torej $Z = 1$; sicer velja $Z = 0$. Nekateri pristopi, osnovani na paradigmi *vreče značilnic* (angl. bag of features), npr. Dorkó in Schmid (2003), takšno enostavno definicijo vrednosti značilnic dejansko uporabljajo. Če pa bi poleg dvojiškega podatka o pojavitvi značilnice upoštevali še podatke o položajih, kjer značilnica Z nastopa, ter o velikostih in orientacijah lokalnih struktur, v katerih se značilnica Z manifestira na sliki, bi opis vrednosti značilnice postal bistveno kompleksnejši.

Na področju vizualne klasifikacije je opis značilnic kot takih na splošno deležen

večjega zanimanja kot opis vrednosti značilnic. Vrednosti značilnic so običajno določene implicitno, kot rezultat reprezentacije slike z značilnicami. Eksplicitno podane vrednosti značilnic so, kot bomo videli kasneje, pomembne le pri uporabi nekaterih izbirnih metod, kot sta medsebojna informacija in mera RELIEF. Sicer je pomembnejši opis samih značilnic, kar pa se v strojnem učenju pogosto abstrahira (zanimaja nas npr. vrednost značilnice „barva“ za nek klasifikacijski primer, ne pa opis samega koncepta barve). Lokalne značilnice za vizualno klasifikacijo so pogosto opisane z numeričnimi vektorji, kar omogoča primerjavo med značilnicami na osnovi razdalje med opisujočimi vektorji. Možnosti pa je seveda še veliko več. Fidler in Leonardis (2007), na primer, opišeta vsako značilnico z rekurzivno podatkovno strukturo, ki odseva hierarhično kompozicijsko zgradbo značilnic.

1.2.2 Izbira značilnic v klasifikacijskih shemah

Klasifikacijski problemi so pogosto predstavljeni z redundantno množico značilnic, zato je vprašanje izbire značilnic na mestu tudi v domeni klasifikacije. Izbira značilnic pripomore k učinkovitejši in točnejši klasifikaciji, poleg tega pa lahko doprinese novo znanje k problemski domeni. Pri medicinski diagnostiki, na primer, je z uporabo izbirnih metod mogoče ugotoviti, kateri simptomi so dejansko relevantni za dano diagnostično domeno.

V teoretskih obravnavah problema izbire značilnic se pogosto uporablja pojem *relevantnosti* značilnic. Obstaja več formalnih definicij relevantnosti (Blum in Langley, 1997; John in sod., 1994; Peng in sod., 2005), vendar jih ne bomo navedli. V nadaljevanju bomo pojem relevantnosti značilnic uporabljali v neformalnem ali intuitivnem smislu kot sopomenko za „primernost“ ali „koristnost“ značilnic za dani klasifikacijski problem.

Po Blumu in Langleyu (1997) je izbirne metode možno uvrstiti v tri skupine:

Vgrajene metode (angl. *embedded methods*): Izbira značilnic je sestavni del postopka za učenje klasifikatorja.

Filtrske metode : Izbira značilnic se izvrši pred izvedbo učnega postopka.

Metode na osnovi notranje optimizacije (angl. *wrappers*): Izbirni postopek sistematično izvaja učni algoritem in preizkuša naučene klasifikatorje za različne kandidatne podmnožice pridobljenih značilnic.

V nadaljevanju bomo navedli nekaj primerov izbirnih pristopov za vsako od naštetih skupin.

1.2.2.1 Vgrajene metode

Pri vgrajenih izbirnih metodah je izbira značilnic neločljivi del samega postopka za učenje klasifikatorja. Primeri vgrajenih pristopov so metode za gradnjo klasifikacijskih dreves, metode za učenje simboličnih pravil, metoda SVM itn. Pri

osnovnem postopku za gradnjo klasifikacijskih dreves (Mitchell, 1997), na primer, se v vsaki iteraciji poišče značilnica, ki omogoča najboljše razločevanje med primeri razredov. V primeru pravočasne ustavitve gradnje drevesa ali v primeru uporabe algoritma naknadnega rezanja bo nastalo drevo vsebovalo le bolj ali manj relevantne značilnice.

V magistrski nalogi se z vgrajenimi pristopi ne bomo ukvarjali. Posvetili se bomo pristopom, pri katerih izbira značilnic nastopa kot samostojen modul in ne kot neločljivi del postopka za učenje klasifikatorja.

1.2.2.2 Filtrske metode

Osrednja značilnost filtrskih metod je izbira značilnic pred učenjem klasifikatorja. V fazi učenja klasifikatorja se tako uporabljajo le izbrane značilnice.

Filtrske metode izbirajo značilnice na osnovi neke *ocenjevalne funkcije*, ki vrednoti koristnost posameznih značilnic za klasifikacijski problem. Pri nekaterih metodah je ocenjevalna funkcija edino merilo za izbiro značilnic. Pri drugih metodah je ocenjevalna funkcija vpeta v kompleksnejši izbirni postopek. Včasih je ocenjevalna funkcija opredeljena implicitno, s samim izbirnim postopkom, in je ni mogoče izluščiti iz izbirnega postopka.

Veliko ocenjevalnih funkcij je zasnovanih za uporabo v dvorazrednih klasifikacijskih problemih, ki so v kontekstu detekcije objektov tudi najbolj zanimivi, saj, kot bomo videli kasneje, je problem detekcije mogoče prevesti na problem dvorazredne klasifikacije. Nekatero od funkcij, ki jih bomo navedli, je mogoče premočrtno razširiti na večrazredno domeno. Ocenjevalne funkcije se razlikujejo tudi po vrstah značilnic, za katere so primerne. Nekatero funkcijo je mogoče uporabiti le z dvojiškimi značilnicami, druge le z diskretnimi, tretje z zveznimi značilnicami, obstajajo pa tudi funkcije, ki so v tem pogledu univerzalne.

Naštejmo nekaj primerov ocenjevalnih funkcij (Dorkó, 2006):

Razmerje verjetij (angl. likelihood ratio). Gre za funkcijo, namenjeno ocenjevanju dvojiških značilnic v domeni dvojiške klasifikacije. Razmerje verjetij za značilnico Z je definirano na sledeči način:

$$L(Z) = \frac{P(Z = 1 | C = 1)}{P(Z = 1 | C = 0)}, \quad (1.1)$$

pri čemer $P(Z = 1 | C = c)$ podaja pogojno verjetnost dogodka, da je vrednost dvojiške značilnice Z enaka 1, če primer pripada pozitivnemu razredu ($c = 1$) oziroma negativnemu razredu ($c = 0$). Razmerje verjetij je asimetrična mera glede na vrednosti značilnice, saj v formuli nastopa le člen ($Z = 1$), ne pa tudi ($Z = 0$). Dorkó in Schmid (2003) sta pokazala primernost mere za uporabo v vizualni klasifikaciji, kjer je člen ($Z = 1$) možno interpretirati kot pojavitev značilnice Z v vizualnem primeru. Po tej interpretaciji je vrednost $L(Z)$ visoka za tiste značilnice, ki se pogosto pojavljajo v pozitivnih, a redko v negativnih primerih. Prav takšne značilnice pa so relevantne za vizualno klasifikacijo.

Medsebojna informacija. Ta funkcija je za razliko od razmerja verjetij simetrična glede na vrednost značilnice Z :

$$I(Z; C) = \sum_{z \in \{0,1\}} \sum_{c \in \{0,1\}} P(Z = z, C = c) \log_2 \frac{P(Z = z, C = c)}{P(Z = z)P(C = c)}. \quad (1.2)$$

Posplošitev na večvrednostne značilnice in večrazredne probleme je očitna. Medsebojna informacija je mera za medsebojno odvisnost med naključnima spremenljivkama. Mera je tem večja, čim bolj zanesljivo je možno na podlagi vrednosti ene naključne spremenljivke (konkretno na podlagi vrednosti značilnice) sklepati na vrednost druge naključne spremenljivke (konkretno na pripadnost razredu). Medsebojno informacijo kot sredstvo za izbiro značilnic med drugim uporabljajo Dorkó in Schmid (2003) ter Epshtein in Ullman (2005).

Točnost pragovnega klasifikatorja. Pri izbiri zveznih ali celoštevilskih diskretnih značilnic je možno uporabiti sledečo hevristiko: koristna je takšna značilnica, ki zavzema razmeroma visoke vrednosti pri pozitivnih primerih in razmeroma nizke vrednosti pri negativnih primerih ali obratno. Značilnica, katere vrednosti so naključno porazdeljene po primerih, je gotovo manj koristna. Koristnost značilnic je možno meriti tako, da se za vsako značilnico zgradi optimalni pragovni klasifikator, nato pa se točnost tega klasifikatorja na učnih primerih obravnava kot ocenjevalna funkcija.

Kot sta pokazala Viola in Jones (2004), je možno tovrstno ocenjevalno funkcijo elegantno vpeti v izbirni postopek AdaBoost, ki ga bomo še večkrat omenili. Zaradi številnih dobrih lastnosti postopka AdaBoost nekatere metode definirajo ocenjevalno funkcijo na osnovi pragovnega klasifikatorja tudi v primeru, ko značilnice niso zvezne. Tak pristop uporabljamo tudi v magistrski nalogi.

Vse tri navedene ocenjevalne funkcije obravnavajo le odvisnosti med posameznimi značilnicami in razrednimi pripadnostmi, *medsebojne odvisnosti značilnic* (Jakulin in Bratko, 2004) pa ne upoštevajo. Vendar se pogosto izkaže, da nabor samih individualno najkoristnejših značilnic vsebuje določeno stopnjo redundance (Cover, 1974), in to prav zaradi odvisnosti med izbranimi značilnicami oziroma, natančneje, zaradi pogojne odvisnosti med izbranimi značilnicami glede na razredno pripadnost. Ilustrirajmo tovrstno odvisnost s preprostim dvojiškim klasifikacijskim problemom. Denimo, da množica značilnic obsega dvojiški značilnici Z_1 in Z_2 , za razredno pripadnost pa velja $C = Z_1 = \overline{Z_2}$. Množica primerov za ta klasifikacijski problem vsebuje le dva različna primera: $Z_1 = 1, Z_2 = 0, C = 1$ in $Z_1 = 0, Z_2 = 1, C = 0$. Funkcija $I(Z; C)$ bi tako za Z_1 kot za Z_2 vrnila največjo možno vrednost ($I(Z_1; C) = I(Z_2; C) = 1$), saj je tako na podlagi vrednosti Z_1 kot tudi na podlagi vrednosti Z_2 moč povsem zanesljivo sklepati na razredno pripadnost C . Vendar pa sta značilnici medsebojno odvisni (tako pogojno glede na razredno pripadnost kot tudi brezpogojno), saj velja $Z_1 = \overline{Z_2}$. Zato bi izbira ene od obeh

značilnic povsem zadoščala za zanesljivo klasifikacijo primerov. Žal pa ocenjevalna funkcija $I(Z; C)$ informacije o stopnji redundantnosti kandidatnih značilnic glede na že izbrane značilnice ne more nuditi.

V nadaljevanju bomo navedli nekaj ocenjevalnih funkcij in izbirnih algoritmov, ki medsebojno odvisnost med značilnicami upoštevajo bodisi eksplicitno bodisi implicitno:

Pogojna medsebojna informacija. Fleuret (2004) je predlagal algoritem, ki na osnovi vrednotenja pogojne medsebojne informacije eksplicitno upošteva odvisnost med značilnicami. Predlagani algoritem v vsaki iteraciji izbere po eno značilnico. V prvi iteraciji izbere značilnico, ki maksimizira medsebojno informacijo značilnica-razred:

$$Z^{[1]} = \arg \max_Z I(Z; C). \quad (1.3)$$

V iteraciji $t > 1$ pa postopek izbere značilnico po naslednjem pravilu:

$$Z^{[t]} = \arg \max_{Z \in \mathcal{Z}'} \left\{ \min_{u < t} I(Z; C | Z^{[u]}) \right\}, \quad (1.4)$$

pri čemer \mathcal{Z}' označuje množico značilnic, ki v prvih $t - 1$ iteracijah še niso bile izbrane, oznaka $Z^{[i]}$ pa predstavlja značilnico, izbrano v i -ti iteraciji.

Pogojno medsebojno informacijo $I(Z; C | Z')$ je možno tolmačiti kot informacijo, ki jo značilnica Z doprinese k poznavanju razredne pripadnosti posameznih primerov (C), ob upoštevanju, da je značilnica Z' že prispevala svoj delež informacije k poznavanju vrednosti C . Če informacijo, ki jo Z nudi za poznavanje spremenljivke C , zajema že značilnica Z' , potem je pogojna informacija $I(Z; C | Z')$ nizka, četudi je brezpogojna informacija $I(Z; C)$ morda visoka. Fleuretov postopek tako v vsaki iteraciji $t > 1$ izbere tisto značilnico, ki glede na že izbrane značilnice doprinese največ informacije k poznavanju razredne pripadnosti učnih primerov.

Izbirno shemo, temelječo na pogojni medsebojni informaciji, uporabljata tudi Epshtein in Ullman (2005). Nekoliko drugačno, vendar prav tako informacijsko-teoretsko utemeljeno izbirno strategijo so zasnovali Peng in sod. (2005).

Algoritem AdaBoost. Ta algoritem sta zasnovala Freund in Schapire (1997) kot sredstvo za kombiniranje preprostih, t.i. šibkih klasifikatorjev z razmeroma nizko klasifikacijsko točnostjo v kompleksnejši, t.i. močni klasifikator z višjo klasifikacijsko točnostjo. AdaBoost v vsaki iteraciji izbere šibki klasifikator z največjo točnostjo na danih učnih primerih glede na uteži, prirejene posameznim primerom. Na začetku postopka so vsi primeri enako uteženi, ob zaključku vsake iteracije pa postopek AdaBoost poveča uteži tistim primerom, ki jih je klasifikator, izbran v dani iteraciji, napačno klasificiral. Sprememba uteži bo vplivala na izbiro klasifikatorja v naslednji iteraciji:

točnost, ki jo bodo kandidatni klasifikatorji dosegli na močnejše uteženih primerih, bo imela večji vpliv kot točnost, ki jo bodo klasifikatorji dosegli na šibkeje uteženih primerih.

Viola in Jones (2004) sta pokazala, da je mogoče postopek AdaBoost uporabiti tudi kot mehanizem za izbiro značilnic, če se vzpostavi pomensko smiselna korespondenca med značilnicami in šibkimi klasifikatorji. Zaradi opisane iterativne uteževalne sheme postopek AdaBoost pri izbiranju značilnic implicitno upošteva odvisnost med značilnicami.

AdaBoost predstavlja ogrodje, v katerega je možno vpeti različne ocenjevalne funkcije. Funkcije, ki same po sebi ne upoštevajo odvisnosti med značilnicami, po vgraditvi v AdaBoost to lastnost samodejno pridobijo. Viola in Jones sta v ogrodje AdaBoost vstavila funkcijo, ki temelji na porazdelitvi vrednosti celoštevilskih značilnic po učnih primerih. Li in sod. (2006) so pokazali, kako je v ogrodje mogoče integrirati funkcijo tujosti, ki jo opisujemo v nadaljevanju.

V novjšem času je AdaBoost priljubljena raziskovalna tema. Poleg del, ki smo jih že ali pa jih še bomo navedli, naštejmo še nekaj pristopov, ki se poslužujejo postopka AdaBoost: Torralba in sod. (2004); Zhang in sod. (2005); Li in Ito (2005); Opelt in sod. (2006a); Grabner in Bischof (2006); Grabner in sod. (2007). Seznam s tem seveda še zdaleč ni izčrpan. Preprosta, učinkovita in neverjetno prilagodljiva zamisel Freunda in Schapireja bo najbrž še nekaj časa predmet zanimanja raziskovalne skupnosti.

Relief. Kira in Rendell (1992) sta definirala mero RELIEF, ki je primerna za ocenjevanje zveznih in diskretnih značilnic v dvorazrednih problemih. Algoritem za izračun mere za vsak učni primer P_i poišče najbližji učni primer iz istega razreda, P'_i , in najbližji učni primer iz nasprotnega razreda, \overline{P}'_i . (Razdalja med danima primeroma je definirana kot razdalja med vektorjema vrednosti vseh značilnic pri danih dveh primerih.) Naj $v(Z, Q)$ označuje vrednost značilnice Z pri nekem primeru Q . Značilnica Z prejme visoko oceno RELIEF, če je razlika $|v(Z, P_i) - v(Z, P'_i)|$ v povprečju (preko vseh i) majhna, razlika $|v(Z, P_i) - v(Z, \overline{P}'_i)|$ pa v povprečju velika. Kononenko (1994) je predlagal nekaj izboljšav osnovne metode, med drugim razširitev na večrazredne klasifikacijske probleme.

Pri metodi RELIEF se odvisnosti med značilnicami implicitno upoštevajo. Razdalja, na podlagi katere se določijo najbližji sosedi obravnavanega primera P_i , je namreč odvisna od vrednosti *vseh* značilnic pri primeru P_i in kandidatih za primere P'_i in \overline{P}'_i .

Tujost (angl. strangeness). (Li in sod., 2006) so definirali mero tujosti za izbiro najkoristnejših slikovnih zaplat za vizualno klasifikacijo. Vsaka zaplata, pridobljena iz neke slike, se obravnava kot samostojna značilnica. Ker slike ustrezajo klasifikacijskim primerom, velja, da vsaka značilnica nastopa v na-

tanko enem primeru. (Nastop značilnice v primeru P bi ustrezal dvojiški vrednosti $Z = 1$ pri primeru P , a to v kontekstu funkcije tujosti ni pomembno.) Pogoji za uporabo mere tujosti je definiranost razdalje med samimi značilnicami oziroma njihovimi opisniki. Razdaljo med zaplatama, na primer, je možno opredeliti kot razdaljo med normaliziranimi vektorjema slikovnih elementov, iz katerih sta zaplati sestavljeni.

Naj bo Z ocenjevana značilnica, pridobljena iz nekega učnega primera razreda c . Naj \mathcal{Z}_c označuje množico vseh značilnic, pridobljenih iz učnih primerov razreda c , množica $\mathcal{Z}_{\bar{c}}$ pa naj vsebuje vse preostale značilnice. Po Li in sod. doseže Z visoko vrednost tujosti, če se glede na definirano mero razdalje nahaja razmeroma daleč od značilnic \mathcal{Z}_c in razmeroma blizu značilnic $\mathcal{Z}_{\bar{c}}$. Takšna značilnica je tujek v svoji okolici, saj so ji bližje značilnice, ki izvirajo iz „tujih“ primerov ($\mathcal{Z}_{\bar{c}}$), kot pa značilnice, ki izvirajo iz „domačih“ primerov (\mathcal{Z}_c). Tujki so najverjetneje irelevantne značilnice, ki bi učenje klasifikatorja kvečjemu ovirale. Zato je smiselno, če izbirni postopek daje prednost značilnicam z nizko stopnjo tujosti.

Kot smo že omenili, so Li in sod. funkcijo tujosti vgradili v izbirni postopek AdaBoost.

1.2.2.3 Metode na osnovi notranje optimizacije

Metode na osnovi notranje optimizacije (Kohavi in John, 1997) uporabljajo postopek za učenje klasifikatorja kot podprogram. V angleščini se takšne metode imenujejo „wrappers“, saj so „ovite“ okrog učnega algoritma. Prevod „notranja optimizacija“ je povzet po Kononenko (2005).

Izbirne metode na osnovi notranje optimizacije praviloma obravnavajo izbiro značilnic kot preiskovanje prostora stanj. Vsako stanje ustreza neki kandidatni množici izbranih značilnic. Izbirni postopek za vsako obiskano stanje izračuna njegovo oceno, in sicer tako, da požene učni algoritem in preizkusi zgrajeni klasifikator na učnih primerih. Učni algoritem mora zgraditi klasifikator na podlagi kandidatne množice značilnic za obiskano stanje. Cilj izbirnega postopka je poiskati najbolje ocenjeno stanje pri danem ciljnem številu izbranih značilnic. Množica stanj je ponavadi prevelika za popoln prebor prostora, zato se velikokrat uporablja kar požrešni način preiskovanja, ki ga usmerjajo ocene obiskanih stanj.

Najbolj pereča slabost metod na osnovi notranje optimizacije je velika računska zahtevnost. Zato se tovrstne metode pogosto uporabljajo le kot dopolnilo osnovnim filtrskim ali vgrajenim izbirnim metodam. Peng in sod. (2005), na primer, uporabijo notranjo optimizacijo le za „fino“ izbiro značilnic; v prvi, „grobi“ fazi uporabijo filtrski pristop, ki temelji na eksplicitnem upoštevanju medsebojne odvisnosti značilnic.

1.2.3 Izbira značilnic v detekcijskih shemah

V tem podrazdelku bomo s študijo primerov prikazali povezavo med domeno (vizualne) klasifikacije in domeno detekcije objektov na slikah. Podrobneje bomo opisali tri sodobne detekcijske metode. Vse tri metode izbirajo značilnice z uporabo postopka AdaBoost, ki zahteva definicijo šibkih klasifikatorjev. Zato vse predstavljene metode prevedejo izvorni problem detekcije objektov v enakovreden problem klasifikacije primerov. Razlikujejo se zlasti po načinu prevedbe in po ocenjevalnih funkcijah, ki jih vstavijo v postopek AdaBoost.

Seveda je značilnice v detekcijskih shemah možno izbirati tudi brez prevedbe detekcije v klasifikacijo. Vendar je v primeru uporabe postopka AdaBoost, ki se je poslužuje tudi izbirna metoda, razvita za magistrsko nalogo, tovrstna prevedba nujna. Poleg tega se, kot smo že omenili, večina obstoječe teorije in prakse izbire značilnic nanaša na klasifikacijske probleme. Zato se s prevedbo detekcije v klasifikacijo odpre možnost uporabe raznovrstnih preizkušenih in teoretsko utemeljenih pristopov k izbiri značilnic v klasifikacijski domeni. Nekaj tovrstnih pristopov smo navedli v podrazdelku 1.2.2.

1.2.3.1 Metoda Viola in Jones (2004)

Zaradi enostavnosti predpostavimo, da so vsi objekti v učnih in testnih slikah prikazani v enaki velikosti, ki je znana že v času učenja. Problem detekcije se tako skrči na iskanje položajev objektov v dani sliki. Po Violi in Jonesu je dvojiški klasifikacijski problem, ki ustreza detekcijskemu, opredeljen s sledečimi postavkami:

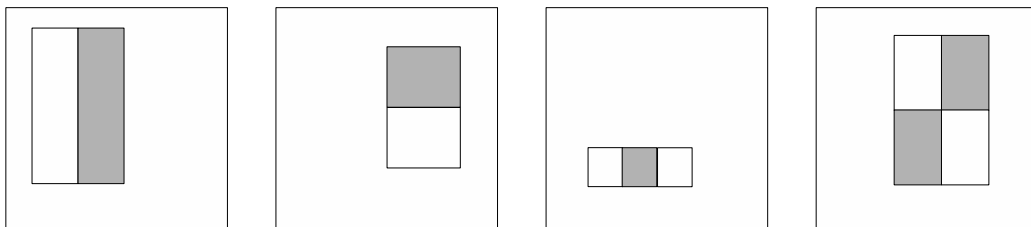
- *Učni primeri* so slikovna okna fiksne velikosti, ki je enaka velikosti objektov. *Pozitiven učni primer* ustreza oknu, ki vsebuje podobo nekega objekta iz dane vizualne kategorije. *Negativen učni primer* je okno, ki lahko vsebuje karkoli razen podobe nekega objekta dane vizualne kategorije.
- Množica *testnih primerov* za dano testno sliko sestoji iz vseh oken, ki jih je mogoče pridobiti iz testne slike. Velikost testnih oken je enaka velikosti učnih oken. Iz slike velikosti $A \times B$ slikovnih elementov je mogoče pridobiti $(A - a + 1)(B - b + 1)$ oken velikosti $a \times b$, če le velja $A \geq a$ in $B \geq b$.
- Naloga *klasifikatorja* je uvrstiti vsako okno, ki izhaja iz dane testne slike, v pozitivni ali negativni razred. Uvrstitev okna v pozitivni razred ustreza detekcijski napovedi, da se na položaju v sliki, iz katerega je bilo izrezano okno, nahaja objekt. Uvrstitev okna v negativni razred napoveduje, da na ustrežajočem položaju objekt ne nastopa. Tako je možno iz napovedi klasifikatorja neposredno pridobiti napovedi položajev objektov v sliki, kar je cilj izvirnega detekcijskega problema.
- *Učni postopek* zgradi klasifikator na podlagi učnih oken.

Tovrstni prehod v klasifikacijsko domeno omogoča uporabo poljubne klasifikacijske izbirne metode, primerne za sestav značilnic, ki služi reprezentiranju slik in

slikovnih oken. V primeru uporabe lokalnih značilnic bi bilo možno uporabiti npr. medsebojno informacijo. Vrednost značilnice Z za dani učni primer (okno) bi bila enaka 1, če značilnica Z v oknu nastopa, sicer pa bi veljalo $Z = 0$. Ker so v učni fazi znane razredne pripadnosti (C) za vsa učna okna, bi bilo možno neposredno izračunati aproksimacije verjetnosti $P(Z = z, C = c)$, $P(Z = z)$ in $P(C = c)$ (za $c \in \{0, 1\}$, $z \in \{0, 1\}$), ki nastopajo v (1.2). Kot bomo videli v nadaljevanju, pa se Viola in Jones izbire značilnic lotita na drugačen način.

Kako je mogoče postopati, če predpostavka o konstantni in znani velikosti testnih objektov ne drži? Po standardni rešitvi, ki se je poslužujeta tudi Viola in Jones, se običajni, enovelikostni testni postopek požene na izvorni testni sliki in na dovolj gosti množici povečanih in pomanjšanih različic testne slike. Tako je mogoče upati, da bo dani testni objekt na eni od različic slike prikazan v velikosti, ki bo dovolj blizu učne velikosti. Množica kopij slik v različnih velikostih je običajno organizirana kot *velikostna piramida* (angl. scale space pyramid, Crowley (1981)).

Viola in Jones uporabljata značilnice s celoštevilskimi vrednostmi. Izračun značilnic je izjemno enostaven in zaradi t.i. integralske slikovne predstavitve, ki sta jo izumila prav Viola in Jones, tudi zelo računsko učinkovit. Vsaka značilnica pripada enemu od štirih tipov, ki so prikazani na sliki 1.3. Pri uporabi oken velikosti 24×24 znaša število vseh pridobljenih značilnic 45 396. Seveda pa se vseh značilnic poslužuje le izbirni postopek; kot sta pokazala Viola in Jones, že dve primerno izbrani in uteženi značilnici zadoščata za presenetljivo dobre rezultate pri detekciji obrazov.



Slika 1.3: Slike prikazujejo štiri tipe vrednotenja značilnic Viola-Jones na danem oknu. Zunanji okvirji ponazarjajo okna, notranji svetli in temni okvirji pa položaje slikovnih elementov, ki sodelujejo pri vrednotenju posameznih značilnic na slikovnem oknu. Vrednost značilnice na oknu je definirana kot razlika med vsoto intenzitet slikovnih elementov v temnem okvirju in vsoto intenzitet v svetlem okvirju. Značilnice lahko nastopajo v poljubnih velikostih in na poljubnih položajih znotraj okna, zato je njihovo število lahko zelo veliko. (Slika je povzeta po Viola in Jones (2004).)

Učna metoda za vsako pridobljeno značilnico definira preprost pragovni klasifikator. Ti klasifikatorji bodo nato nastopali kot šibki klasifikatorji v postopku AdaBoost. Ker klasifikatorji ustrezajo posameznim značilnicam, bo izbira klasifikatorjev v postopku AdaBoost dejansko enakovredna izbiri značilnic. Vsak šibki klasifikator sestoji iz značilnice, naučenega praga in naučenega predznaka. Op-

timalne vrednosti pragov in predznakov so odvisne od uteži učnih oken, zato se določijo v vsaki iteraciji postopka AdaBoost posebej.

Šibki klasifikator uvrsti učno okno v pozitivni razred natanko tedaj, ko je vrednost njegove značilnice na oknu večja (ali manjša, odvisno od naučenega predznaka) od naučenega praga. V vsaki iteraciji postopka AdaBoost se prag in predznak klasifikatorja nastavita tako, da klasifikator doseže največjo možno točnost na učnih oknih glede na trenutne uteži oken. Klasifikatorji za značilnice, ki zavzemajo visoke vrednosti na pozitivnih oknih in nizke vrednosti na negativnih oknih (ali obratno), bodo vsaj v prvi iteraciji postopka AdaBoost, ko so vsa okna enakovredno utežena, dosegli visoko stopnjo točnosti. Če pa je vrednost značilnice bolj ali manj naključno porazdeljena po učnih oknih, bo pripadajoči klasifikator tudi pri optimalnem pragu in predznaku dosegel razmeroma nizko točnost.

V postopku AdaBoost se v vsaki iteraciji izbere najuspešnejši šibki klasifikator glede na trenutne vrednosti uteži učnih oken. Hkrati s klasifikatorjem se seveda izbere tudi pripadajoča značilnica. Postopek AdaBoost ob koncu združi izbrane šibke klasifikatorje v močni pragovni klasifikator, ki dosega bistveno boljše rezultate na učnih oknih (in, tako je mogoče upati, tudi na testnih oknih) kot posamezni šibki klasifikatorji.

Viola in Jones sta v istem članku predstavila še eno inovacijo. Gre za t.i. *pozornostno kaskado* močnih klasifikatorjev. Testna okna se najprej obdelajo z močnim klasifikatorjem, ki zajema zelo malo značilnic (morda dve ali tri). Tak klasifikator nemara ni najbolj zanesljiv, vendar je njegovo obnašanje mogoče uravnavati s pragom. Če je prag nastavljen dovolj nizko, bo klasifikator pozitivna okna skoraj zanesljivo proglasil za pozitivna, vendar pa bo v pozitivni razred uvrstil tudi veliko negativnih oken. Na podlagi pozitivne klasifikacije danega okna zato ni mogoče zanesljivo sklepati na dejansko razredno pripadnost okna. Povsem drugače pa je z negativno klasifikacijo. Zaradi nizko postavljenega praga bo negativna uvrstitev okna (v nasprotju s pozitivno uvrstitvijo) razmeroma redek, a tudi zanesljiv dogodek. Okna, ki jih bo nizkopragovni klasifikator proglasil za negativna, bodo torej skoraj zanesljivo negativna. Zato je možno negativno klasificirana okna izločiti iz nadaljnje obravnave, saj skoraj gotovo ne vsebujejo objekta. Pozitivno klasificirana okna pa se obdelajo z naslednjim klasifikatorjem v kaskadi, ki zajema več značilnic kot prvi klasifikator. Ker je večina oken v testnih slikah negativnih, je pri kaskadni obdelavi možno že z zelo enostavnim klasifikatorjem izločiti velik del slike. Zato metoda Viole in Jonesa ni uspešna zgolj v smislu dosežene detekcijske točnosti, temveč tudi v smislu časovne učinkovitosti. Kot poročata Viola in Jones, je njuna metoda za obdelavo slike ločljivosti 384×288 slikovnih elementov pri vseh velikostih v velikostni piramidi potrebovala zgolj 67 milisekund na računalniku s 700-megahercnim procesorjem Intel Pentium III.

Morda se bo bralcu zazdelo, da smo metodi Viole in Jonesa namenili nesorazmerno veliko prostora. Vendar menimo, da si ta prostor glede na svojo pionirskost in vplivnost tudi zasluži. Metoda namreč predstavlja navdih za celo vrsto znanih in manj znanih metod, razvitih v zadnjih nekaj letih. Kot smo omenili, se tudi naše magistrsko delo napaja iz spoznanj Viole in Jonesa.

1.2.3.2 Metoda Shotton in sod. (2005)

Shotton in sod. so se prevedbe detekcije v klasifikacijo lotili na drugačen način. Medtem ko Viola in Jones klasificirata posamezna okna, pridobljena iz testne slike, Shotton in sod. klasificirajo posamezne položaje v testni sliki. Klasifikatorjeva uvrstitev položaja v pozitivni razred ustreza detekcijski napovedi, da se na danem položaju v testni sliki nahaja središče (težišče) nekega objekta fiksne in znane velikosti.

Metoda Shottona in sod. uporablja t.i. obrisne (konturne) značilnice, ki so zmožne zajeti zgolj informacijo o obliki objektov dane kategorije, teksturne informacije pa ne morejo upoštevati. Poglavitni cilj metode je bil namreč po navedbah Shottona in sod. izpostaviti pomembnost oblike pri vizualni klasifikaciji oz. detekciji. Značilnice so pridobljene neposredno iz robov segmentacijskih mask učnih slik. (Segmentacijska maska je dvojiška matrika, ki podaja, kateri slikovni elementi pripadajo objektu, kateri pa ozadju.) Vsaka značilnica, Z , je opisana z več parametri. Eden od parametrov, \mathbf{p} , je vektorska razlika med položajem značilnice na segmentacijski maski in položajem središča objekta na ustrezajoči učni sliki.

Klasifikator sklepa takole: Denimo, da se središče nekega objekta nahaja na položaju \mathbf{c} na testni sliki. Če to drži, je precej verjetno, da v okolici položaja $\mathbf{c} + \mathbf{p}$ nastopa značilnica Z . To velja za vse pridobljene značilnice in pripadajoče parametre \mathbf{p} . Klasifikator nato izračuna odstopanje med dejanskimi položaji značilnic in napovedanimi položaji $\mathbf{c} + \mathbf{p}$. Čim manjše je to odstopanje, tem večja je verjetnost, da bo klasifikator točko \mathbf{c} proglasil za pozitivno. Klasifikator seveda ne pozna dejanskih položajev objektov v testni sliki, zato mora ovrednotiti vse slikovne koordinate na dovolj gosti mreži.

Klasifikator, ki se uporablja v testni fazi, je kombinacija šibkih klasifikatorjev, določena z uporabo postopka AdaBoost. Vsak šibki klasifikator je podobno kot pri metodi Viola-Jones definiran z značilnico in optimalnim pragom, zato AdaBoost dejansko tudi izbira značilnice, ne samo kombinira šibke klasifikatorje. Učenje šibkih klasifikatorjev poteka na osnovi množice pozitivnih in negativnih učnih primerov. Učni primeri so slikovni položaji na učnih slikah. Položaji v bližini središča nekega učnega objekta se obravnavajo kot pozitivni učni primeri, ostali položaji pa kot negativni učni primeri.

Shotton in sod. postopka AdaBoost ne uporabljajo zgolj za izbiro značilnic in gradnjo končnega klasifikatorja, temveč tudi za učenje dveh parametrov samih značilnic. AdaBoost v vsaki iteraciji izbere en element iz trirazsežne strukture. V eni razsežnosti so nanizani šibki klasifikatorji oz. značilnice, vsaka od preostalih dveh razsežnosti pa vsebuje diskretizirano množico vrednosti ustreznega parametra. Po trditvah avtorjev gre za izviren pristop k uporabi algoritma AdaBoost.

1.2.3.3 Metoda Opelt in sod. (2006c)

Opelt in sod. uporabljajo enak sestav značilnic kot Shotton in sod. Medtem ko metoda Shottona in sod. na podlagi predpostavljenega (testna faza) ali znanega

(učna faza) položaja središča objekta sklepa na položaje posameznih značilnic, Opelt in sod. uberejo obratno pot: na podlagi položajev nastopov značilnic sklepajo na položaj središča nekega objekta v dani sliki. Tovrsten način sklepanja je sicer osrednja značilnost priljubljene detekcijske paradigme, ki so jo razvili Leibe in sod. (2004).

Pri Opeltu in sod. so učni primeri kar posamezne učne slike ali njihovi deli. Podobe objektov so pozitivni, podobe ozadja pa negativni učni primeri. Podobno kot Viola in Jones ter Shotton in sod. tudi Opelt in sod. definirajo šibke klasifikatorje, ki jih nato izbirajo v postopku AdaBoost. Vendar pa je šibki klasifikator po metodi Opelta in sod. odvisen od skupine k značilnic, ne zgolj od ene značilnice. (Opelt in sod. uporabljajo $k = 2$ ali $k = 3$.) Razlog za takšno odločitev je povečati točnost šibkih klasifikatorjev, ki vstopajo v AdaBoost. Ker je število vseh možnih skupin k značilnic pogosto preveliko, jih metoda v fazi „predizbire“ izbere po določenih kriterijih.

Šibki klasifikator je definiran tako, da je njegova točnost skladna s koristnostjo dane skupine značilnic za detekcijo učnih slik. Skupina značilnic je intuitivno koristna, če se pogosto pojavlja na pozitivnih učnih slikah in redko na negativnih slikah ter če je na podlagi značilnic iz skupine dovolj pogosto možno pravilno sklepati na položaj središča objekta v učnih slikah. Za praktično uporabo je bilo omenjeni kriterij treba formalizirati v obliki neke funkcije. Šibki klasifikator določi razredno pripadnost dani učni sliki glede na vrednost te funkcije za dano skupino značilnic na učni sliki in glede na vrednost naučenega optimalnega praga. Tako je vse pripravljeno za uporabo postopka AdaBoost, ki zgradi končni močni klasifikator in hkrati izbere relevantne skupine značilnic.

Omeniti velja še članek, ki so ga isti avtorji objavili na konferenci CVPR 2006. Ta članek (Opelt in sod., 2006b) predstavlja inkrementalno metodo izbire značilnic in učenja močnega klasifikatorja. Naučeni klasifikator ni uporabljen zgolj za detekcijo objektov posameznih vizualnih kategorij, temveč tudi za določanje neznane kategorije prikazanega objekta.

1.2.3.4 Končne opombe

Na podlagi predstavljenih metod je bralec morda dobil občutek, da vse sodobne izbirne metode za vizualno detekcijo temeljijo na prevedbi detekcije v klasifikacijo in uporabi postopka AdaBoost. To ne drži. Res pa je, da je metoda AdaBoost učinkovita, uspešna, prilagodljiva in preprosta, zato jo je smiselno izkoristiti, če ta možnost obstaja.

Eno od detekcijskih metod, ki sloni na prevedbi detekcije v klasifikacijo, vendar ne uporablja postopka AdaBoost, so zasnovali Agarwal in sod. (2004). Podobno kot Viola in Jones tudi Agarwal in sod. detektirajo objekte na slikah preko klasificiranja slikovnih oken. Za izbiro značilnic in za učenje klasifikatorja pa Agarwal in sod. uporabljajo učno arhitekturo, imenovano Sparse Network of Winnows.

Z ozirom na poudarjanje pretvorbe detekcije v klasifikacijo se zdi, kot da je tovrstna pretvorba edina možnost za izbiro značilnic za vizualno detekcijo. To

seveda ne drži. Prav lahko si je zamisliti kakšno *ad hoc* metodo ali ocenjevalno funkcijo, ki vrednoti značilnice brez definicije klasifikatorja. Opelt in sod., na primer, definirajo klasifikator zgolj zato, da lahko funkcijo za ocenjevanje značilnic vpnejo v postopek AdaBoost. Seveda pa bi lahko značilnice izbirali zgolj na podlagi ocenjevalne funkcije, v testni fazi pa bi preprosto izvajali detekcijski postopek na osnovi izbranih značilnic. Klasifikatorjem bi se tako povsem izognili.

Uvrščanje predstavljenih metod glede na shemo izbiranja značilnic ni povsem enoznačno. Postopek AdaBoost bi lahko prištevali med vgrajene izbirne metode, saj izbira značilnic poteka hkrati z gradnjo močnega klasifikatorja. Vendar pa osnovo za AdaBoost vselej predstavlja neka ocenjevalna funkcija, ki bi načeloma že sama po sebi mogla služiti namenu izbire značilnic. Zaradi tega bomo predstavljene metode uvrstili v filtrsko skupino.

Pri razvrščanju izbirnih metod za klasifikacijsko domeno smo omenili tudi metode na osnovi notranje optimizacije. Prilagoditev tovrstnih metod za detekcijsko domeno je povsem naravna. Namesto da bi izbirni postopek pri preiskovanju prostora kandidatnih množic izbranih značilnic gradil in preizkušal klasifikatorje, gradi in preizkuša detekcijske modele. Cilj izbirnega postopka je poiskati najboljše ocenjeno množico izbranih značilnic v danem prostoru stanj, pri čemer je ocena stanja definirana kot točnost, ki jo detekcijski model, zgrajen na podlagi množice značilnic za to stanje, doseže na učni množici.

1.3 Magistrska naloga

1.3.1 Pregled

V magistrski nalogi predstavljamo dve metodi za izbiro značilnic v detekcijski shemi; prva deluje po filtrskem pristopu, druga pa po pristopu na osnovi notranje optimizacije.

Filtrska izbirna metoda, ki smo jo razvili, se v precejšnji meri zgleduje po metodah Viola in Jones (2004) in Opelt in sod. (2006b), vendar uporablja izvirno funkcijo za ocenjevanje značilnic. Izbirna metoda ni vezana na specifičen detekcijski pristop. Prav tako metoda ne omejuje sestava lokalnih značilnic, ki se uporablja za reprezentacijo slik.

Razvili smo tudi izbirno metodo na osnovi notranje optimizacije. Metoda temelji na splošnem načelu snovanja tovrstnih metod in na klasičnem požrešnem pristopu k preiskovanju prostora stanj, zato ji ne pripisujemo izvirnosti. Kljub temu pa metoda po našem mnenju predstavlja obogatitev magistrske naloge, saj gre za način izbire značilnic, ki na področju vizualne detekcije in klasifikacije ni posebej uveljavljen. Načelo „ovitja“ izbirnega postopka okrog detekcijskega modula zagotavlja, da je metodo na osnovi notranje optimizacije mogoče kombinirati s poljubno detekcijsko metodo.

V magistrski nalogi smo obe razviti izbirni metodi povezali z detekcijsko metodo, ki temelji na pristopu Leibe in sod. (2004). Preizkusili smo dve različni metodi za pridobivanje lokalnih značilnic iz slik: metodo Fidler in sod. (2006) in

metodo Ferrari in sod. (2006a). Vsako izmed štirih kombinacij izbirnih metod in sestavov značilnic smo preizkusili na petih različnih vizualnih kategorijah in dobljene detekcijske rezultate ovrednotili v skladu s standardi, ki so jih vzpostavili Agarwal in sod. (2004) ter Leibe (2004). Ugotovili smo, da je v nekaterih primerih že z zelo majhnim številom izbranih značilnic moč doseči presenetljivo visoko detekcijsko točnost. Zasluge za vzpodbudne rezultate je mogoče pripisati vsem trem poglavitnim komponentam celotnega sistema: sestavu značilnic, izbirni metodi in detekcijski metodi.

V magistrski nalogi podajamo izčrpen opis obeh izbirnih metod, detekcijske metode in obeh metod za pridobivanje značilnic. Pri opisovanju detekcijskega modula namenjamo poseben poudarek večvelikostni detekciji. Predstavljamo dva različna pristopa k detekciji objektov pri poljubni velikosti.

1.3.2 Pričakovani prispevek k znanosti

Izpostavili bomo sledeča prispevka magistrske naloge k znanosti:

- Predlog izvirne funkcije za ocenjevanje značilnic za vizualno detekcijo in integracija predlagane funkcije v izbirni postopek AdaBoost.
- Povezava predstavljenih izbirnih metod s sodobnim pristopom k vizualni detekciji in z dvema sodobnima metodama za pridobivanje značilnic iz slik.

Nekaj lastnih zamisli smo vgradili tudi v detekcijsko metodo in v obe metodi za pridobivanje značilnic. Vendar te zamisli praviloma zadevajo tehnične podrobnosti opisanih metod in temeljijo na znanih in preizkušanih rešitvah, zato jih ne želimo izpostavljati kot izvirne prispevke k znanosti.

1.3.3 Implementacijske opombe

Skoraj celoten sistem smo realizirali z lastno implementacijo. Edino izjemo predstavlja del metode za pridobivanje značilnic po Ferrariju in sod., kjer smo uporabili prosto dostopno implementacijo. Naša implementacija detekcijskega modula in metod za pridobivanje značilnic v večji meri sledi postopkom, objavljenim v izvornih člankih, ponekod pa smo udeležili lastne tehnične zamisli.

Sistem smo skoraj v celoti implementirali v okolju Matlab®. Za dosego čim večje časovne učinkovitosti smo se posluževali vektorizacijskih prijemov povsod, kjer je bilo to smiselno. Nekatere kritične programske odseke detekcijskega postopka, ki jih ni bilo mogoče učinkovito vektorizirati, smo realizirali v jeziku C++ in jih preko Matlabove tehnologije MEX integrirali v Matlabove funkcije. Časovna učinkovitost detekcijskega postopka je med drugim ključni dejavnik pri metodi na osnovi notranje optimizacije.

1.4 Zgradba nadaljnjega besedila

Poglavje 2 namenjamu podrobnejšemu opisu obeh implementiranih metod za izbiro značilnic, filterske in notranjeoptimizacijske. V poglavju 3 je poglobljeno opisan celoten detekcijski modul, ki zajema postopek za učenje modela, detekcijski postopek in modul za vrednotenje rezultatov detekcije. Poseben poudarek namenjamu problemu detekcije objektov pri poljubni velikosti. V poglavju 4 sta predstavljena oba uporabljena sestava značilnic. Pri obeh sestavih opisujemo zgradbo značilnic, postopek za pridobivanje značilnic in postopek za reprezentacijo slik s pridobljenimi značilnicami. Poglavje 5 je namenjeno prikazu in vrednotenju rezultatov sistema na petih različnih vizualnih kategorijah. Poglavje 6 sklene magistrsko nalogo.

Izbira značilnic

2.1 Uvod

V tem poglavju podrobneje opisujemo oba pristopa k izbiri značilnic za vizualno detekcijo, ki smo ju implementirali in preizkusili. V okviru razvijanja filtrskega pristopa smo definirali ocenjevalno funkcijo na osnovi prevedbe problema detekcije v ustrezajoč problem klasifikacije. Ocenjevalno funkcijo smo vgradili v izbirno ogrodje AdaBoost. Notranjeoptimizacijski pristop obravnava problem izbire značilnic kot problem preiskovanja prostora stanj.

Oba izbirna pristopa uporabljata rezultate detekcijskega modula nad dano učno slikovno množico. Vendar pa detekcijski modul v veliki meri obravnavata kot „črno škatlo“, s katero izbirni postopek komunicira le preko natančno določenih vhodov in izhodov. Zato je oba izbirna pristopa mogoče kombinirati z raznovrstnimi detekcijskimi metodami, ne zgolj z metodo, ki jo opisujemo v poglavju 3. Za združljivost z izbirnimi metodama mora detekcijski modul izpolnjevati le pogoje glede pričakovanega vhoda in izhoda.

Pri opisovanju obeh izbirnih metod se bomo posluževali sledeče notacije. Naj $\mathcal{Z} = \{Z_1, \dots, Z_N\}$ označuje neko izhodiščno množico značilnic. V okviru magistrske naloge gre povsod za množico, ki jo postopek za pridobivanje značilnic ustvari na podlagi neke učne slikovne množice za dano kategorijo. V izbirni postopek vstopata množica \mathcal{Z} in zeleno število izbranih značilnic, ki ga bomo označevali s T . Izbirni postopek vrne množico T izbranih značilnic, $\mathcal{Z}_T = \{Z^{[1]}, \dots, Z^{[T]}\}$. Množico \mathcal{Z}_T bomo imenovali tudi *izbirna množica*. Obe izbirni metodi izbirata značilnice v iterativni shemi, zato za vsak par (T, T') , tako da je $T < T'$, velja $\mathcal{Z}_T \subset \mathcal{Z}_{T'}$. Izbirni postopek dejansko vrne *zaporedje* izbranih značilnic, ki se z večanjem parametra T zgolj podaljšuje.

Predstavimo še zgradbo pričujočega poglavja. V razdelku 2.2 navajamo predpostavke glede vhoda in izhoda detekcijskega modula. Predlagano filtrsko metodo za izbiro značilnic opisujemo v razdelku 2.3, metodo na osnovi notranje optimizacije pa v razdelku 2.4.

2.2 Pričakovani vhod in izhod detekcijskega modula

Oba izbirna pristopa predpostavljata, da detekcijski modul na vhodu sprejme naslednje podatke:

- Poljubno množico značilnic $\mathcal{Z}' \subseteq \mathcal{Z}$.
- Množico slik. Izbira značilnic se izvaja v učni fazi, zato detekcijski modul tekom izbirnega postopka obdeluje posebej pripravljeno učno slikovno množico, ki je na sliki 1.2 označena kot „Učne slike II“ in ki jo bomo podrobneje opisali v odseku 2.3.1.1.

Pri iskanju detekcijskih hipotez se detekcijski modul sme posluževati le značilnic iz \mathcal{Z}' in manifestacij teh značilnic v podanih učnih slikah.

Detekcijski postopek po predpostavki zgradi seznam hipotez o položajih in velikostih objektov dane kategorije v učnih slikah. Vsaka hipoteza izraža domnevo o nastopanju nekega objekta na določeni sliki. Filtrska izbirna metoda predpostavlja, da detekcijski modul za vsako detekcijsko hipotezo H priskrbi naslednje podatke:

- Nenegativno realno število $s(H)$, imenovano *podpora* ali *moč* hipoteze, ki podaja detektorjevo stopnjo zaupanja v pravilnost hipoteze H . Visoka vrednost $s(H)$ nakazuje, da detekcijski postopek močno „verjame“, da je hipoteza pravilna, in obratno. Seveda pa vrednost $s(H)$ ne pove ničesar o *dejanski* pravilnosti hipoteze.
- Odločitev o tem, ali je hipoteza H dejansko pravilna.

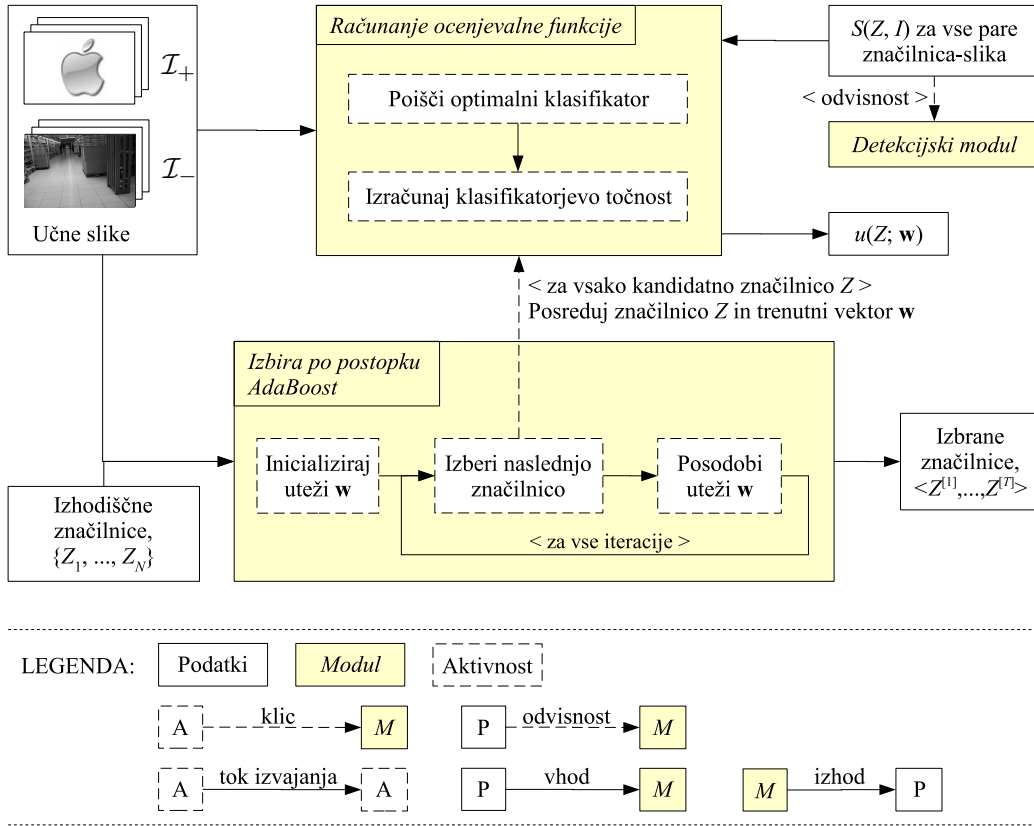
V nadaljnjem besedilu bo izraz „najmočnejša hipoteza“ označeval detekcijsko hipotezo z največjo vrednostjo $s(H)$.

Metoda za notranjeoptimizacijsko izbiro značilnic podatkov o posameznih hipotezah ne potrebuje. Potrebuje le podatek o *detekcijski točnosti*, ki jo detekcijski postopek doseže na celotni učni množici z uporabo podane množice značilnic. Računanje točnosti (razdelek 3.5) seveda temelji na preverjanju pravilnosti hipotez in na vrednostih $s(H)$, ki jih proizvede detekcijski postopek, vendar se izbirni metodi po načelu „črne škatle“ z vrednotenjem detekcijske točnosti ni treba ukvarjati. Za vse potrebno poskrbi detekcijski modul.

2.3 Filtrska metoda za izbiro značilnic

Predstavitev filtrske izbirne metode pričenjamo z opisom prevedbe problema detekcije v problem klasifikacije (podrazdelek 2.3.1). Podrazdelek 2.3.2 je posvečen postopni izpeljavi ocenjevalne funkcije, ki temelji na omenjeni prevedbi. V podrazdelku 2.3.3 predstavljamo vključitev ocenjevalne funkcije v izbirno ogrodje AdaBoost.

Celoten postopek izbire ilustrira slika 2.1.



Slika 2.1: Shematski prikaz predlagane metode za izbiro značilnic po filtrskem pristopu. Izbirni postopek AdaBoost v vsaki iteraciji izbere po eno značilnico ter posodobi vektor slikovnih uteži \mathbf{w} . Pri izbiranju se vsaka kandidatna značilnica ovrednoti z ocenjevalno funkcijo, ki poleg značilnice prejme tudi vektor \mathbf{w} . Vrednosti $S(Z, I)$ so odvisne od rezultatov detekcijskega algoritma na učnih slikah. Ker pa so neodvisne od vektorja \mathbf{w} , se lahko izračunajo vnaprej. (Navedeni koncepti so obrazloženi v nadaljnjem besedilu.)

2.3.1 Prevedba problema detekcije v problem klasifikacije

2.3.1.1 Klasifikacijska razreda in klasifikacijski primeri

Definicija ustreznega problema klasifikacije se prične z opredelitvijo obeh klasifikacijskih razredov in njihovih primerov. Oba razreda, *pozitivni* in *negativni*, vsebujeta učne slike. Vsaka učna slika tako predstavlja primer iz enega od obeh razredov. Slike iz pozitivnega razreda bomo imenovali kar „pozitivne slike“, slike iz negativnega razreda pa „negativne slike“.

Za slike v pozitivnem razredu, ki ga bomo označevali z \mathcal{I}_+ , morajo veljati sledeče lastnosti:

- Vsaka slika v \mathcal{I}_+ prikazuje natanko en objekt dane vizualne kategorije.

- Vsaki sliki je pridružena informacija o *omejevalnem okvirju* vsebovanega objekta. V celotnem izbirnem postopku se uporabljajo zgolj deli slik znotraj omejevalnih okvirjev.

Ker klasifikacija učnih slik vključuje uporabo detekcijskega postopka nad slikami, so dobrodošle, četudi ne obvezne, lastnosti, ki detekcijskemu postopku olajšajo delo. Koristno je, na primer, če so objekti prikazani na enotnem (npr. belem) ozadju ali pa so kako drugače ločeni od ozadja.

Slike v negativnem razredu, \mathcal{I}_- , smejo prikazovati karkoli *razen* objektov dane kategorije. Sicer za \mathcal{I}_- ne velja nobena druga omejitev. V primeru, ko nas zanima le ena kategorija (in je vse ostalo „ozadje“), je koristno, če negativni razred vsebuje čimbolj raznovrstne slike. V primeru, ko želimo izbiro značilnic eksplicitno usmeriti proti neki alternativni kategoriji (npr. kategorija avtomobilov proti kategoriji tovornjakov), pa je smiselno, če negativni razred vsebuje slike te alternativne kategorije. V konkretnem primeru bi pozitivni razred vseboval podobe avtomobilov, negativni pa (med drugim) podobe tovornjakov.

Izbirni algoritem na svojem vhodu pričakuje označene slike iz razredov \mathcal{I}_+ in \mathcal{I}_- , ki zadoščajo navedenim pogojem. Predpostavljali bomo nepraznost obeh razredov. Celotno učno množico bomo označevali z \mathcal{I} . Velja $\mathcal{I} = \mathcal{I}_+ \cup \mathcal{I}_-$.

2.3.1.2 Klasifikacijski problem

Klasifikacijski problem je definiran kot uvrščanje slik iz \mathcal{I} v pozitivni ali negativni razred na podlagi klasifikatorja, ki ga bomo definirali v nadaljevanju. Klasifikator posredno uporablja rezultate detekcijskega postopka nad slikami iz \mathcal{I} , vendar bomo ta vidik izpostavili šele v naslednjem odseku, kjer bo definirana tudi funkcija za ocenjevanje značilnic.

Naj Z označuje neko značilnico iz \mathcal{Z} . Definirajmo preprost pragovni klasifikator, ki uvršča učne slike v razrede po sledečem pravilu:

$$d_{Z,\theta}(I) = \begin{cases} 1, & \text{če } S(Z, I) \geq \theta; \\ 0, & \text{če } S(Z, I) < \theta. \end{cases} \quad (2.1)$$

Funkcija $S : \mathcal{Z} \times \mathcal{I} \rightarrow \mathbb{R}$ uporablja rezultate detekcijskega algoritma in bo definirana v naslednjem odseku. Zaenkrat ji ne bomo pripisovali nobenega določnega pomena.

Klasifikator (2.1) uvrsti učno sliko I v pozitivni razred ($d_{Z,\theta}(I) = 1$), če vrednost funkcije S za značilnico Z in sliko I doseže ali preseže prag θ . V nasprotnem primeru klasifikator označi sliko kot pripadnico negativnega razreda. Klasifikator $d_{Z,\theta}$ je definiran po vzoru šibkega klasifikatorja v metodi Viola in Jones (2004).

Klasifikator razvršča slike na podlagi funkcije S in praga θ , *dejanske* razredne pripadnosti pa seveda ne pozna. Naj y_I označuje dejanski razred slike I :

$$y_I = \begin{cases} 1, & \text{če } I \in \mathcal{I}_+; \\ 0, & \text{če } I \in \mathcal{I}_-. \end{cases} \quad (2.2)$$

Definirajmo *stopnjo klasifikacijske napake* klasifikatorja $d_{Z,\theta}$:

$$\epsilon(Z, \theta; \mathbf{w}) = \sum_{I \in \mathcal{I}} w_I |d_{Z,\theta}(I) - y_I|, \quad (2.3)$$

kjer je $\mathbf{w} = (w_I)_{I \in \mathcal{I}}$ vektor uteži in

$$w_I = \begin{cases} 1/(2|\mathcal{I}_+|), & \text{če } I \in \mathcal{I}_+; \\ 1/(2|\mathcal{I}_-|), & \text{če } I \in \mathcal{I}_-. \end{cases} \quad (2.4)$$

predstavlja uteži posameznih slik. Uteži so določene tako, da je skupna teža pozitivnih slik enaka skupni teži negativnih slik, in sicer $1/2$. Takšna definicija omogoča enostavno vključitev funkcije za ocenjevanje značilnic v postopek Ada-Boost. Od podrazdelka 2.3.3 naprej bomo vektor uteži, definiran z (2.4), raje označevali kot $\mathbf{w}^{[1]}$, saj bo oznaka \mathbf{w} nastopala kot spremenljivka in ne kot na vrednost $\mathbf{w}^{[1]}$ postavljena konstanta. Do takrat pa bo poenostavitev $\mathbf{w} \equiv \mathbf{w}^{[1]}$ priročna in upravičena.

Izraz $|d_{Z,\theta}(I) - y_I|$ v (2.3) znaša 0 za pravilno klasificirane slike in 1 za napačno klasificirane slike. Stopnja klasifikacijske napake je tako omejena na interval $[0, 1]$. Uvedimo še pojem *klasifikacijske točnosti* in ga definirajmo kot $(1 - \epsilon(Z, \theta; \mathbf{w}))$.

Pri podani značilnici Z je stopnja napake klasifikatorja $d_{Z,\theta}$ močno odvisna od praga θ . *Optimalni prag* za značilnico Z je definiran kot prag, pri katerem stopnja napake doseže najnižjo možno vrednost:

$$\theta^*(Z; \mathbf{w}) = \arg \min_{\theta} \epsilon(Z, \theta; \mathbf{w}). \quad (2.5)$$

Optimalni klasifikator za značilnico Z je definiran kot klasifikator za Z , ki uporablja optimalni prag: $d_Z^* \equiv d_{Z,\theta^*(Z;\mathbf{w})}$. Ta klasifikator doseže med vsemi klasifikatorji za Z najnižjo stopnjo napake. Njegova stopnja napake je odvisna od porazdelitve vrednosti funkcije $S(Z, I)$ preko celotne množice \mathcal{I} , ta pa je odvisna od značilnice Z . Ne oziraje se na porazdelitev $(S(Z, I))_{I \in \mathcal{I}}$ pa velja, da je stopnja napake optimalnega klasifikatorja vselej manjša ali enaka 0,5. Po formuli (2.3) bi namreč tolikšno stopnjo napake ne glede na porazdelitev vrednosti $S(Z, I)$ dosegel klasifikator $d_{Z,-\infty}$, ki vsako sliko uvrsti v pozitivni razred. Stopnja napake optimalnega klasifikatorja je po definiciji kvečjemu manjša.

Za določitev optimalnega praga za dano značilnico Z zadošča pregled največ $|\mathcal{I}| + 1$ kandidatov. To trditev bomo na kratko utemeljili. Naj $\langle S_1, \dots, S_{|\mathcal{I}|} \rangle$ predstavlja naraščajoče urejeno zaporedje vrednosti $\{S(Z, I)\}_{I \in \mathcal{I}}$. Klasifikator (2.1) pri poljubnem pragu $\theta \in (-\infty, S_1]$ uvrsti vse slike v pozitivni razred. Pri kateremkoli $\theta \in (S_i, S_{i+1}]$ (za $i = 1, \dots, |\mathcal{I}| - 1$) so slike I z lastnostjo $S(Z, I) \leq S_i$ uvrščene v negativni, vse ostale slike pa v pozitivni razred. Pri poljubnem $\theta \in (S_{|\mathcal{I}|}, \infty)$ klasifikator označi vse slike kot negativne. Število različnih možnih uvrstitev in s tem število različnih vrednosti (2.3) znaša kvečjemu $|\mathcal{I}| + 1$, kot kandidati za optimalni prag pa lahko nastopajo katerekoli vrednosti iz posameznih pragovnih intervalov.

2.3.1.3 Pregled

Zaključimo razdelek s pregledom elementov, ki določajo prevedbo problema detekcije v problem dvojiške klasifikacije:

Definicija razredov in primerkov razredov: Primerki obeh razredov, pozitivnega in negativnega, so posamezne učne slike. Pozitivne slike vsebujejo podobe objektov in so označene z omejevalnimi okvirji podob. Negativne slike takih podob ne smejo vsebovati.

Klasifikator: Klasifikator $d_{Z,\theta}$ razvršča učne slike na osnovi primerjanja vrednosti še ne definirane funkcije $S : \mathcal{Z} \times \mathcal{I} \rightarrow \mathbb{R}$ s pragom θ .

Ocenjevanje klasifikatorja: Mera za kakovost klasifikatorja je njegova točnost ($= 1 - \text{stopnja napake}$), ki temelji na primerjanju klasifikatorjevih napovedi in dejanskih razredov za posamezne učne slike. Optimalni klasifikator doseže izmed vseh klasifikatorjev $d_{Z,\theta}$ pri fiksnem $Z \in \mathcal{Z}$ in prostem $\theta \in \mathbb{R}$ najvišjo točnost.

Povezava z izbiro značilnic: Vez med klasifikacijskim problemom in izbiro značilnic predstavlja funkcija S , ki je odvisna od podane značilnice in ki služi kot osnova za klasifikacijo.

Povezava z detekcijo: Navezava na detekcijski algoritem je prav tako domena funkcije S .

V naslednjem podrazdelku bomo definirali funkcijo za ocenjevanje značilnic in funkcijo S ter tako podrobneje razdelali zadnji dve postavki.

2.3.2 Funkcija za ocenjevanje značilnic

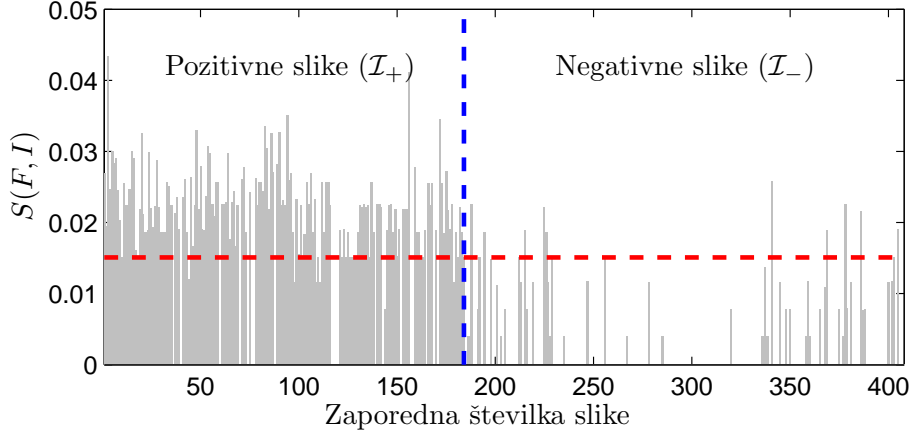
Ocenjevalna funkcija vrednoti potencialno koristnost posameznih značilnic za detekcijo. Definirana je kot klasifikacijska točnost optimalnega klasifikatorja za značilnico Z :

$$u(Z; \mathbf{w}) = 1 - \epsilon(Z, \theta^*(Z; \mathbf{w}); \mathbf{w}). \quad (2.6)$$

Ocenjevalna funkcija je seveda odvisna od še ne opredeljene funkcije $S : \mathcal{Z} \times \mathcal{I} \rightarrow \mathbb{R}$. Funkcija S mora biti definirana tako, da bo osmislila ocenjevalno funkcijo; kot smo navedli ob koncu prejšnjega razdelka, funkcija S predstavlja vez med klasifikacijskim problemom, izbiro značilnic in detekcijo. Pred opredelitvijo funkcije S pa bomo opisali obnašanje ocenjevalne funkcije glede na porazdelitev $(S(Z, I))_{I \in \mathcal{I}}$ pri fiksni značilnici Z .

Denimo, da funkcija $S(Z, I)$ pri dani značilnici konsistentno zavzema relativno visoke vrednosti pri pozitivnih slikah in relativno nizke vrednosti pri negativnih slikah. V tem primeru bo optimalni prag ležal nekje na sredini med visokimi in nizkimi vrednostmi, optimalni klasifikator pa bo pravilno napovedal razredno pripadnost za večino učnih slik. Zaradi nizke stopnje napake bo značilnica Z v

skladu z definicijo (2.6) prejela visoko oceno. Porazdelitev $(S(Z, I))_{I \in \mathcal{I}}$ za razmeroma visoko ocenjeno značilnico je prikazana na sliki 2.2.



Slika 2.2: Porazdelitev $(S(Z, I))_{I \in \mathcal{I}}$ za najbolje ocenjeno značilnico za kategorijo skodelic pri uporabi značilnic Fidler in sod. (razdelek 4.2). Vodoravna črtkana črta prikazuje optimalni prag. Navpična črtkana črta je ločilo med pozitivnimi in negativnimi slikami. Optimalni klasifikator napačno uvrsti 26 od 184 pozitivnih slik in 16 od 224 negativnih slik. Stopnja klasifikacijske napake tako znaša $16/(2 \times 224) + 26/(2 \times 184) = 0,106$, ocena značilnice pa je enaka $1 - 0,106 = 0,894$.

Če pa bi vrednosti $S(Z, I)$ bile porazdeljene enakomerno ali naključno po slikovni množici ali če bi funkcija S zavzemala relativno visoke vrednosti za negativne in relativno nizke vrednosti za pozitivne slike, potem niti optimalni klasifikator ne bi bil zmožen doseči nizke stopnje napake. Zaradi neugodne porazdelitve $(S(Z, I))_{I \in \mathcal{I}}$ bi katerakoli vrednost praga θ vodila do velikega števila napačnih klasifikacij. Posledično bi ocenjevalna funkcija dodelila značilnici Z nizko oceno.

Funkcijo S je najbolje definirati tako, da bo formalna definicija ocenjevalne funkcije ustrezala naši intuitivni predstavi o potencialni koristnosti dane značilnice za detekcijo. Po intuitivni definiciji je značilnica Z potencialno koristna, če velja sledeče:

- Značilnica Z se pogosto pojavlja v pozitivnih in redko v negativnih učnih slikah. Gre za lastnost *distinktivnosti*.
- Če je detekcijski postopek omejen le na uporabo značilnice Z (kar pomeni, da mora tvoriti detekcijske hipoteze zgolj na podlagi manifestacij značilnice Z v učnih slikah), potem bo pri večini pozitivnih slik tvoril pravilne detekcijske hipoteze. Značilnica s takšno lastnostjo je *prediktivna*.

Naj omenimo, da distinktivnost in prediktivnost nista nujno povezani lastnosti. Značilnica je lahko distinktivna, a neprediktivna (v pozitivnih slikah se sicer pojavlja bistveno pogosteje kot v negativnih, vendar njene pojavitve ne nudijo dovolj

informacije za tvorbo točnih hipotez), lahko pa je tudi prediktivna, a nedistinktivna (omogoča tvorbo dobrih hipotez v pozitivnih slikah, vendar pogosto nastopa tudi v negativnih slikah). Visoko oceno potencialne koristnosti si zasluži le značilnica, ki je distinktivna in hkrati prediktivna.

Kot bomo pokazali, sledeča definicija funkcije S omogoča, da ocenjevalna funkcija u vrednoti potencialno detekcijsko koristnost dane značilnice v skladu z našo intuitivno predstavo:

- Za poljubno pozitivno sliko I_+ :

$$S(Z, I_+) = \begin{cases} s(H_{\max}(I_+) | Z), & \text{če je } H_{\max}(I_+) \text{ pravilna;} \\ 0 & \text{v nasprotnem primeru.} \end{cases} \quad (2.7)$$

- Za poljubno negativno sliko I_- :

$$S(Z, I_-) = s(H_{\max}(I_-) | Z). \quad (2.8)$$

$s(H_{\max}(I) | Z)$ označuje moč najmočnejše detekcijske hipoteze za sliko I pri pogoju, da detekcijski algoritem pri gradnji detekcijskih hipotez uporablja zgolj manifestacije značilnice Z . Pravilnost hipoteze se določi s posebnim postopkom v okviru detekcijskega modula (razdelek 3.5).

Če je dana značilnica Z potencialno koristna v intuitivnem smislu, pogosto nastopa v pozitivnih in redko v negativnih slikah ter omogoča detekcijskemu algoritmu, da postavi mnogo pravih hipotez. Če predpostavimo, da hipoteze črpajo svojo moč iz pojavitev značilnic v sliki, potem bo takšna značilnica najverjetneje povzročila, da bodo vrednosti $S(Z, I)$ razmeroma visoke za pozitivne in razmeroma nizke za negativne slike. Posledica tega bo nizka stopnja napake optimalnega klasifikatorja in visoka ocena značilnice.

Če značilnica Z ni potencialno koristna v intuitivnem pomenu, je nedistinktivna, neprediktivna ali pa oboje. Če je nedistinktivna, je porazdelitev vrednosti $S(Z, I)$ preko slikovne množice bolj ali manj naključna, zato bo celo optimalni klasifikacijski prag vodil do visoke stopnje napake. Ocenjevalna funkcija u bo značilnici Z dodelila nizko oceno. Če je značilnica neprediktivna, bo v skladu z (2.7) vrednost $S(Z, I)$ enaka nič za veliko pozitivnih slik. V tem primeru bodo ničelne vrednosti preprečile, da bi optimalni klasifikator dosegel nizko stopnjo napake, zato bo ocena neprediktivne značilnice Z nizka. Značilnica, ki ni potencialno koristna v intuitivnem smislu, torej ne more doseči visoke ocene.

Po zaslugi ustrezne definicije funkcije S je ocenjevalna funkcija u konsistenten model intuitivne definicije značilnične potencialne koristnosti za reševanje dane detekcijske naloge.

2.3.3 Vgradnja ocenjevalne funkcije v izbirni postopek AdaBoost

Najenostavnejši izbirni postopek izbira značilnice kar po padajočih vrednostih ocenjevalne funkcije. Pri podanem ciljnem številu značilnic T bi tak postopek

torej vrnil množico T najbolj ocenjenih značilnic. Vendar pa, kot smo razložili že v uvodnem poglavju, ta pristop ni najboljši, saj zanemarja medsebojno odvisnost med značilnicami. Pogosto se namreč zgodi, da se med najbolj ocenjenimi značilnicami znajde precej takšnih značilnic, ki so si vizualno podobne in se tudi v slikah pojavljajo na podobnih mestih. Izbirne množice, pridobljene na osnovi padajočih ocen, zato pogosto vsebujejo precej redundance.

O postopku AdaBoost, ki medsebojno odvisnost značilnic upošteva na impliciten način, smo marsikaj povedali že v uvodnem poglavju. Na tem mestu se osredotočamo na tehnično izvedbo prilagoditve tega postopka za domeno izbire značilnic.

Ocenjevalna funkcija (2.6) je bila zasnovana tako, da omogoča elegantno vključitev v ogrodje AdaBoost. Pri definiciji ocenjevalne funkcije nastopajo uteži w_I , ki podajajo prispevek posameznih slik k stopnji napake optimalnega klasifikatorja. Postopek AdaBoost, ki je prikazan kot algoritem 2.1, uporablja in iterativno spreminja te uteži. V vsaki iteraciji postopek brez vračanja izbire eno značilnico iz množice Z .

Na začetku postopka AdaBoost se uteži učnih slik nastavijo tako, kot narekuje zapis (2.4). AdaBoost nato v vsaki iteraciji določi optimalni klasifikator za vsako še ne izbrano značilnico. Postopek izbire tisto značilnico, pri kateri optimalni klasifikator doseže najnižjo stopnjo napake. Ob koncu vsake iteracije se zmanjšajo uteži tistim slikam, ki jih je optimalni klasifikator za izbrano značilnico pravilno klasificiral (vrstica 7 v algoritmu 2.1, primer $d^{[t]}(I) = y_I$). Slike, pri katerih je klasifikator pogrešil, bodo tako v naslednji iteraciji nastopale z večjo težo. Sprememba uteži bo vplivala na izračun stopnje napake in s tem tudi na določitev optimalnih klasifikatorjev ter izbiro značilnice. Točnost, ki jo kandidatni klasifikatorji dosežejo na močnejše uteženih slikah, je namreč pomembnejša od točnosti, dosežene na šibkeje uteženih slikah. Postopek AdaBoost zato vzpodbuja popravilo napak predhodno izbranih klasifikatorjev. Posledično se zaporedne izbire klasifikatorjev medsebojno dopolnjujejo. Ker med klasifikatorji in značilnicami velja bijektivna preslikava, se medsebojno dopolnjujejo tudi zaporedne izbire značilnic. Posledica takega mehanizma je nižja stopnja redundance v nastali izbirni množici.

V poglavju 5 bomo med drugim primerjali detekcijsko sposobnost množic značilnic, izbranih s postopkom AdaBoost, in množic, izbranih po padajočih ocenah. V kontekstu izbire po padajočih ocenah je ocena značilnice definirana kot $u(Z; \mathbf{w}^{[1]})$, kjer $\mathbf{w}^{[1]}$ podaja *izhodiščne* vrednosti uteži — te so določene z (2.4). Vendar ocenjevalna funkcija nastopa tudi v postopku AdaBoost. V tem primeru se njena vrednost izračuna kot $u(Z; \mathbf{w}^{[t]})$, kjer $\mathbf{w}^{[t]}$ podaja vektor uteži *v iteraciji* t . Če se postopek AdaBoost ne uporabi, se ocena dane značilnice torej ovrednoti glede na uteži $\mathbf{w}^{[1]}$, sicer pa je ocena odvisna od AdaBoostove iteracije in se ovrednoti glede na uteži $\mathbf{w}^{[t]}$. Vrednost $u(Z; \mathbf{w}^{[1]})$ bomo imenovali *statična ocena*, vrednost $u(Z; \mathbf{w}^{[t]})$ pa *dinamična ocena*.

Pridevnika „statično“ in „dinamično“ bomo zaradi enostavnejšega izražanja razširili tudi na druge predstavljene koncepte. „Statično“ se bo vselej nanašalo na uteži $\mathbf{w}^{[1]}$, „dinamično“ pa na uteži $\mathbf{w}^{[t]}$. Statični optimalni klasifikator za

Algoritem 2.1 Izbirni postopek AdaBoost.

Vhod:

- (a) Množica učnih slik \mathcal{I} z oznakami razredne pripadnosti y_I , definiranimi s predpisom (2.2).
- (b) Množica značilnic \mathcal{Z} .
- (c) Želeno število iteracij, T .

Izhod: Zaporedje izbranih značilnic $\langle Z^{[1]}, \dots, Z^{[T]} \rangle$.

Postopek:

- 1: $\mathcal{Z}' := \mathcal{Z}$; // \mathcal{Z}' : množica še ne izbranih značilnic
- 2: Izračunaj izhodiščne vrednosti uteži w_I po formuli (2.4). Naj bo $w_I^{[1]} := w_I$ za vsak $I \in \mathcal{I}$.
- 3: **za** $t = 1$ **do** T **ponavljaj**
- 4: Normaliziraj uteži tako, da njihova vsota postane enaka 1:

$$w_I^{[t]} := \frac{w_I^{[t]}}{\sum_{J \in \mathcal{I}} w_J^{[t]}} \text{ za vse } I \in \mathcal{I}. \quad (2.9)$$

- 5: Izberi značilnico z najvišjo oceno glede na *trenutne* vrednosti uteži:

$$Z^{[t]} = \arg \max_{Z \in \mathcal{Z}'} u(Z; \mathbf{w}^{[t]}). \quad (2.10)$$

Naj $d^{[t]}$ označuje optimalni klasifikator za izbrano značilnico $Z^{[t]}$, $\theta^{[t]}$ prag tega klasifikatorja, $\epsilon^{[t]}$ pa njegovo napako:

$$\begin{aligned} \theta^{[t]} &= \theta^*(Z^{[t]}; \mathbf{w}^{[t]}), \\ d^{[t]} &= d_{Z^{[t]}, \theta^{[t]}}, \\ \epsilon^{[t]} &= \epsilon(Z^{[t]}, \theta^{[t]}; \mathbf{w}^{[t]}) = 1 - u(Z^{[t]}; \mathbf{w}^{[t]}). \end{aligned} \quad (2.11)$$

- 6: $\mathcal{Z}' := \mathcal{Z}' \setminus \{Z^{[t]}\}$.
- 7: Posodobi uteži:

$$w_I^{[t+1]} = \begin{cases} \frac{\epsilon^{[t]}}{1 - \epsilon^{[t]}} w_I^{[t]}, & \text{če } d^{[t]}(I) = y_I; \\ w_I^{[t]}, & \text{v nasprotnem primeru.} \end{cases} \quad (2.12)$$

- 8: **konec** {**za**}
-

značilnico Z doseže najnižjo stopnjo napake glede na uteži $\mathbf{w}^{[1]}$, dinamični pa glede na uteži $\mathbf{w}^{[t]}$. Pri fiksnem naboru učnih slik je stopnja napake statičnega optimalnega klasifikatorja odvisna le od podane značilnice, stopnja napake dinamičnega optimalnega klasifikatorja pa je odvisna tudi od obravnavane iteracije postopka AdaBoost.

2.4 Notranjeoptimizacijska metoda za izbiro značilnic

2.4.1 Uvod

Notranjeoptimizacijski izbirni postopek sistematično poganja detekcijski algoritem na fiksni množici učnih slik (\mathcal{I}) z uporabo različnih kandidatnih množic značilnic. Za vsako izvedbo detekcijskega poskusa se zabeleži dobljena detekcijska točnost. Izbirni postopek skuša poiskati takšno kandidatno množico, ki pri danem ciljnim številu značilnic T omogoča doseg najvišje detekcijske točnosti na slikovni množici \mathcal{I} . Množica \mathcal{I} , za katero ne veljajo nikakršne formalne omejitve, se v primeru notranjeoptimizacijske izbire obravnava kot nedeljiva celota in ne kot unija tujih si množic \mathcal{I}_+ in \mathcal{I}_- .

Problem iskanja optimalne kandidatne množice značilnic je možno naravno formulirati v domeni preiskovanja prostora stanj. V podrazdelku 2.4.2 predstavljamo prostor stanj, mero za ocenjevanje posameznih stanj in usmerjen graf, ki modelira prostor stanj. Podrazdelek 2.4.3 opiše dva uporabna algoritma za iskanje (sub)optimalnega stanja glede na ocenjevalno mero. Graf prostora stanj je ne glede na uporabljen algoritem možno preiskovati v dveh diametralno nasprotnih smereh. To je tema podrazdelka 2.4.4. S podrazdelkom 2.4.5 zaključimo obravnavo notranjeoptimizacijske izbire.

Izbiri značilnic na osnovi notranje optimizacije ilustrira slika 2.3.

2.4.2 Prostor stanj

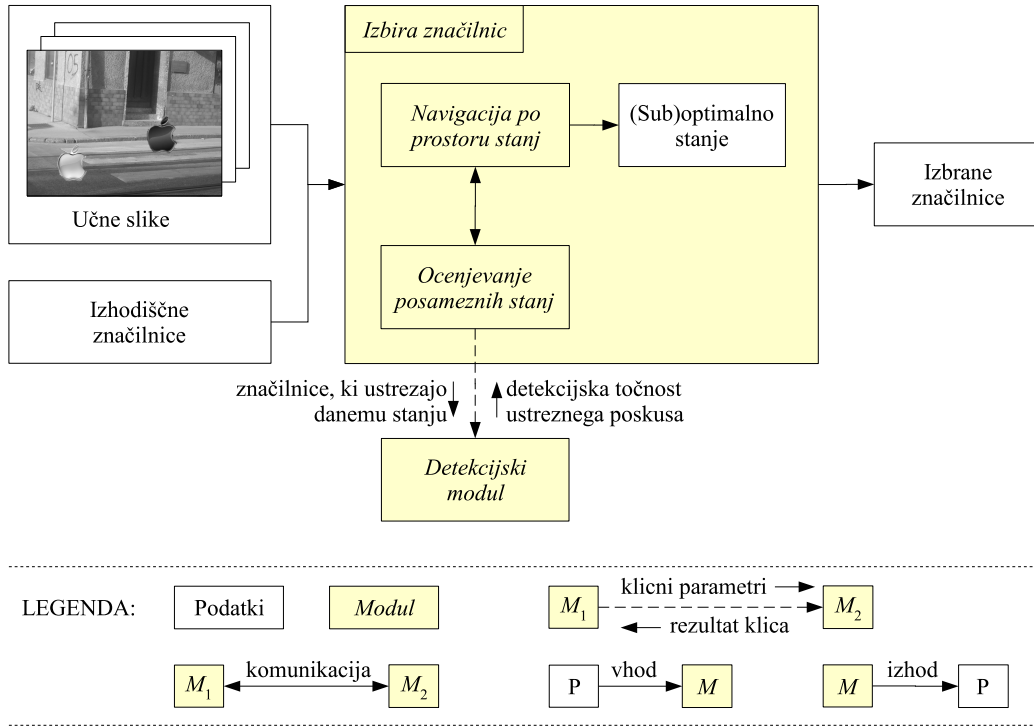
Vsako *stanje* ustreza neki podmnožici izhodiščne množice značilnic \mathcal{Z} . Stanja bomo označevali kar z množicami indeksov pripadajočih značilnic. Stanje $\{J(1), \dots, J(k)\}$ tako ustreza množici $\{Z_{J(1)}, \dots, Z_{J(k)}\} \subseteq \mathcal{Z}$, pri čemer $J : \mathbb{N} \rightarrow \mathbb{N}$ označuje indeksno funkcijo.

Ocena stanja je detekcijska točnost, ki jo detekcijski algoritem doseže na množici učnih slik, pri čemer sme za tvorbo hipotez uporabljati zgolj manifestacije značilnic iz množice, ki jo zajema stanje. Smiselno je, da je merilo za izračun detekcijske točnosti podobno tistemu, ki se bo uporabljalo v testni fazi.

Prostor stanj je mogoče obravnavati kot usmerjen graf, v katerem stanja nastopajo kot vozlišča, množica naslednikov vozlišča (stanja) $V = \{J(1), \dots, J(k)\}$ pa je definirana na sledeč način:

$$\mathcal{N}(V) = \{\{J(1), \dots, J(k), r\} \mid (Z_r \in \mathcal{Z}) \wedge (r \notin V)\} \quad (2.13)$$

Naslednik danega stanja V s k značilnicami je torej vsako stanje s $k + 1$ značilnicami, ki poleg vseh značilnic stanja V vsebuje še natanko eno značilnico, ki še ni vključena v V . Ker je graf možno na zelo naraven način razviti v enakovredno drevo, je smiselno uvesti pojme „koren“, „list“ in „globina“ grafa, tako da le-ti pomensko ustrezajo pojmom v enakovrednem drevesu. V korenu grafa nastopa stanje brez značilnic, označeno z \emptyset . V edinem listu grafa nastopa stanje $\{1, \dots, |\mathcal{Z}|\}$. (Enakovredno drevo bi na ravni listov vsebovalo $|\mathcal{Z}|!$ kopij tega

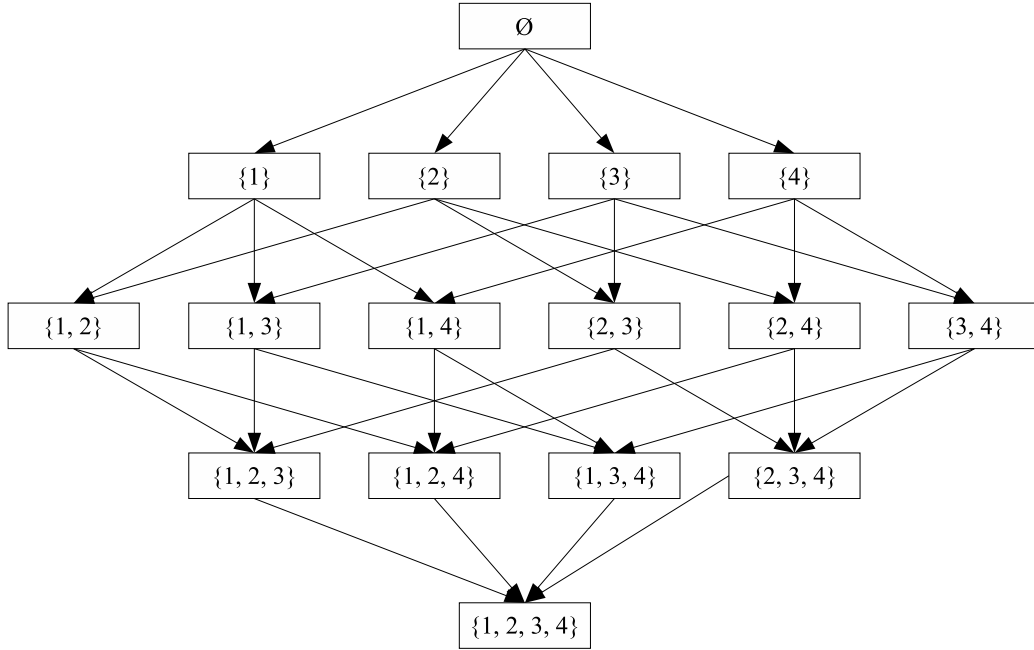


Slika 2.3: Izbira značilnic na osnovi notranje optimizacije. Modul za izbiro značilnic si je možno predstavljati kot skupek dveh med seboj prepletenih modulov. Modul za navigacijo po prostoru stanj preiskuje prostor stanj in kot rezultat vrne (sub)optimalno stanje. Preiskovanje prostora je krmiljeno z ocenami obiskanih stanj. Modul za ocenjevanje stanj za svoje delovanje potrebuje detekcijski modul.

vozljuča.) Na globini d v grafu se nahaja $\binom{|\mathcal{Z}|}{d}$ stanj z d značilnicami. Slika 2.4 predstavlja celoten graf za $|\mathcal{Z}| = 4$.

2.4.3 Iskalne strategije

Naj T podaja (največje) želeno število izbranih značilnic. Cilj izbirnega postopka je poiskati najvišje ocenjeno stanje v prostoru, ki vsebuje stanja za vse podmnožice \mathcal{Z} z največ T značilnicami. Število kandidatnih stanj znaša $\sum_{i=1}^T \binom{|\mathcal{Z}|}{i}$ in naglo narašča s povečevanjem množice \mathcal{Z} in vrednosti T : pri $|\mathcal{Z}| = 100$ in $T = 10$ je število stanj enako že $1,9 \times 10^{13}$. Zaradi tega izčrpno preiskovanje prostora stanj za množice realnih velikosti ni izvedljivo. Kot edina možnost se zato ponuja uporaba približnih, *suboptimalnih* preiskovalnih postopkov. Predstavili bomo požrešni in snopovni preiskovalni postopek.

Slika 2.4: Celoten graf prostora stanj za primer $|\mathcal{Z}| = 4$.

2.4.3.1 Požrešno iskanje

Požrešni postopek preiskovanja prostora stanj (angl. greedy search) je prikazan kot algoritem 2.2. Postopek prične z iskanjem v korenu grafa stanj. V vsaki iteraciji oceni vse naslednike trenutnega vozlišča in se preprosto pomakne v najboljše ocenjenega naslednika. Postopek se ustavi na globini T oziroma tedaj, ko so vsi nasledniki kvečjemu slabši od trenutno obravnavanega vozlišča. Kot smo navedli, izračun ocene vozlišča zahteva izvedbo detekcijskega poskusa na slikovni množici \mathcal{I} .

2.4.3.2 Snopovno iskanje

Nekoliko robustnejša posplošitev požrešnega iskanja je t.i. snopovno iskanje (angl. beam search), ki je prikazano kot algoritem 2.3. Postopek v vsaki iteraciji ohrani M do sedaj najboljše ocenjenih vozlišč, ne zgolj eno samo. Trenutna množica najboljše ocenjenih vozlišč sestavlja *snop*, število M pa podaja njegovo *širino*. Snop v naslednji iteraciji sestavlja M najboljših vozlišč iz unije trenutnega snopa in množice vseh naslednikov vozlišč iz trenutnega snopa. Postopek se ustavi, ko je število značilnic v najboljše ocenjenem vozlišču enako T ali pa ko trenutnega snopa ni več mogoče izboljšati.

Algoritem 2.2 Izbira značilnic s požrešnim preiskovanjem prostora stanj.

Vhod:

- (a) Množica značilnic \mathcal{Z} .
- (b) Množica učnih slik \mathcal{I} .
- (c) Maksimalno ciljno število izbranih značilnic, T .

Izhod: Množica izbranih značilnic.

Postopek:

- 1: $V := \emptyset$; // trenutno vozlišče
 - 2: $d := 0$;
 - 3: $V' := \arg \max_{W \in \mathcal{N}(V)} \text{ocena}(W)$; // najboljši naslednik
 - 4: **dokler** $(d < T) \wedge (\text{ocena}(V) < \text{ocena}(V'))$ **ponavlja**
 - 5: $V := V'$;
 - 6: $V' := \arg \max_{W \in \mathcal{N}(V)} \text{ocena}(W)$;
 - 7: $d := d + 1$
 - 8: **konec** {**dokler**};
 - 9: **vrni** množico značilnic vozlišča V
-

Algoritem 2.3 Izbira značilnic s snopovnim preiskovanjem prostora stanj.

Vhod:

- (a) Množica značilnic \mathcal{Z} .
- (b) Množica učnih slik \mathcal{I} .
- (c) Maksimalno ciljno število izbranih značilnic, T .
- (d) Širina snopa, M .

Izhod: Množica izbranih značilnic.

Postopek:

- 1: $\mathcal{S} := \{\emptyset\}$; // trenutni snop
 - 2: $\mathcal{S}' :=$ množica M najbolj ocenjenih vozlišč iz $\mathcal{N}(\emptyset)$; // naslednji snop
 - 3: $V := \arg \max_{W \in \mathcal{S}'} \text{ocena}(W)$;
 - 4: **dokler** $(|V| < T) \wedge (\mathcal{S}' \neq \mathcal{S})$ **ponavlja**
 - 5: $\mathcal{S} := \mathcal{S}'$;
 - 6: $\mathcal{A} := \emptyset$;
 - 7: **za** $i := 1$ **do** M **ponavlja**
 - 8: $\mathcal{A} := \mathcal{A} \cup \mathcal{N}(\mathcal{S}[i])$ // $\mathcal{S}[i]$ je i -ti element (vozlišče) množice \mathcal{S}
 - 9: **konec** {**za**};
 - 10: $\mathcal{S}' :=$ množica M najbolj ocenjenih vozlišč iz \mathcal{A} ;
 - 11: $V := \arg \max_{W \in \mathcal{S}'} \text{ocena}(W)$
 - 12: **konec** {**dokler**};
 - 13: **vrni** množico značilnic vozlišča V
-

2.4.3.3 Druge možnosti

Za iskanje optimalnega stanja bi načeloma mogli uporabiti poljuben iskalni algoritem. Kot smo že omenili, izčrpne preiskovalne metode, kot so iskanje v globino,

iskanje v širino in iterativno poglobljanje, zaradi obsežnosti prostora stanj v praksi niso izvedljive. V poštev pa pridejo pristopi na osnovi simuliranega ohlajanja (Kirkpatrick in sod., 1983) in genetskih algoritmov (Goldberg, 1989), pa tudi nekatere druge metode.

2.4.4 Smer preiskovanja grafa stanj

Graf prostora stanj je možno preiskovati v dveh smereh: od korena proti listu ali od lista proti korenu. Prva možnost se imenuje *inkrementalno izbiranje* (angl. forward selection), druga pa *vzratna eliminacija* (angl. backward elimination) (Kohavi in John, 1997). Predstavljena iskalna postopka obiskujeta vozlišča grafa stanj v smeri inkrementalnega izbiranja, saj pričneta v stanju \emptyset in napredujeta v smeri proti končnemu stanju $\{1, \dots, |\mathcal{Z}|\}$. Z zamenjavo začetnega in končnega stanja in z obratom funkcije naslednika pa bi predstavljena postopka preiskovala graf stanj v smeri vzratne eliminacije.

Kot argumentirata Kohavi in John, je vzratna eliminacija tipično sposobna odkriti manj redundantno množico značilnic kot inkrementalno izbiranje. Poglavitni problem te strategije pa je velika poraba virov, saj sta časovna in prostorska zahtevnost detekcijskega postopka običajno vsaj linearno odvisni od števila obravnavanih značilnic. Pri nekaj sto kandidatnih značilnicah postane praktična uporabnost te metode v kontekstu detekcije objektov močno vprašljiva.

Zanimiva pa je kombinacija obeh pristopov. Denimo, da preiskovanje prostora stanj v obeh smereh poteka na požrešen način. V prvi fazi bi se izvajalo inkrementalno izbiranje, dokler bi se ocene stanj povečevale oziroma dokler ne bi število izbranih značilnic doseglo neke zgornje meje. Druga faza bi vršila vzratno eliminacijo s pričetkom v vozlišču, v katerem se je inkrementalno izbiranje izteklo. Pri $|\mathcal{Z}| = 4$ bi možno zaporedje izbranih kandidatnih množic izgledalo npr. takole: $\emptyset \rightarrow \{3\} \rightarrow \{3, 1\} \rightarrow \{3, 1, 4\}$ (konec inkrementalnega izbiranja) $\rightarrow \{1, 4\}$. Vozlišče $\{1, 4\}$ je ocenjeno vsaj tako dobro kot $\{3, 1, 4\}$, drugače ga požrešna vzratna eliminacija ne bi odkrila. Inkrementalno požrešno izbiranje pa je to vozlišče zgrešilo, ker najboljše vozlišče na globini 1 ustreza množici $\{3\}$, zaradi česar vsa nadaljnja vozlišča, obiskana v smeri inkrementalnega izbiranja, vsebujejo indeks 3. Tako je na preprost način mogoče izboljšati rezultat običajnega iskalnega postopka, ki napreduje od korena proti listu.

2.4.5 Končne opombe

V implementiranem sistemu se je izkazalo, da je poraba časa izredno omejujoč dejavnik pri notranjeoptimizacijski izbiri značilnic. Morali smo se sprijazniti z določenimi neugodnimi kompromisi, če smo želeli zagotoviti časovno izvedljivost preizkusov. Zaradi tega v poglavju 5 navajamo le rezultate preizkusov požrešnega postopka v smeri inkrementalnega izbiranja. Vse druge možnosti so se žal pokazale kot časovno neizvedljive ali pa (zaradi uvedenih kompromisov) kot nesmiselne.

2.5 Povzetek poglavja

V tem poglavju smo predstavili dva bistveno različna pristopa k izbiranju značilnic. Filtrski pristop temelji na definiciji klasifikatorja nad učnimi slikami, ki deluje kot vez med izbiro značilnic in detekcijo. Klasifikator, povezan z neko značilnico Z , razvršča učne slike v pozitivni in negativni razred glede na vrednosti funkcije S , ki je definirana na osnovi rezultatov detekcije objektov v učnih slikah z uporabo značilnice Z . Ocena značilnice se nato izračuna kot klasifikacijska točnost klasifikatorja, ki je odvisna od porazdelitve vrednosti funkcije S po učnih slikah, ta pa je pogojena s potencialno koristnostjo ocenjevanje značilnice za detekcijo objektov dane kategorije. Postopek za ocenjevanje značilnic je mogoče vključiti v algoritem AdaBoost, ki pri izbiranju značilnic implicitno upošteva medsebojno odvisnost med značilnicami.

Notranjeoptimizacijski pristop k izbiranju značilnic temelji na preiskovanju prostora stanj, čigar elementi — stanja — ustrezajo posameznim kandidatnim množicam izbranih značilnic. Vsakemu stanju je pripisana ocena, ki je definirana kot točnost odgovarjajočega detekcijskega poskusa. Pri izvajanju tega poskusa je detekcijski postopek omejen na uporabo množice značilnic, ki ustreza ocenjevanemu stanju. Cilj izbirnega postopka je poiskati najboljše ocenjeno stanje. Prostor stanj je možno delno urediti z operatorjem naslednika, kar omogoča uporabo raznovrstnih iskalnih algoritmov. Predstavili smo dva suboptimalna, a učinkovita iskalna pristopa, in sicer požrešni in snopovni algoritem. Preiskovanje grafa stanj je mogoče izvajati v smeri od korena proti listu (inkrementalno izbiranje) ali v smeri od lista proti korenu (vzratna eliminacija). Vzratna eliminacija lahko zmanjša redundanco v izbirni množici, pridobljeni z inkrementalnim izbiranjem.

Detekcijski modul

3.1 Uvod

Metoda za učenje in uporabo detekcijskega modela temelji na pristopu Leibe in sod. (2004), ki je podrobneje predstavljen v Leibe (2004). Pristop zgradi in uporabi t.i. implicitni model oblike (angl. Implicit Shape Model; kratica ISM), ki je precej odporen na šum, zakrivanja in variabilnost znotraj kategorije. ISM služi kot osnova za številne druge uspešne detekcijske in klasifikacijske metode (Fritz in sod., 2005; Mikolajczyk in sod., 2006; Thomas in sod., 2006).

ISM upošteva tako informacijo o nastopanju posameznih značilnic v slikah dane kategorije kot tudi informacijo o geometrijski (prostorski) razporeditvi značilnic znotraj posameznih slik. Po stopnji upoštevanja geometrijske informacije je ISM možno uvrstiti med model vreče značilnic (angl. bag-of-features model) in konstelacijski model. Pristop na osnovi modela vreče značilnic (Csurka in sod., 2004; Jurie in Triggs, 2005) upošteva zgolj podatke o pojavitvah posameznih značilnic v danih slikah, podatke o geometrijski konstelaciji značilnic znotraj slik pa povsem zanemari. Tak model zajame le toliko informacije o obliki, kolikor je zajamejo same lokalne značilnice. Model je sicer računsko učinkovit, vendar ni primeren za detekcijo kategorij, pri katerih so objekti opredeljeni predvsem s svojo obliko. V nasprotju z njim skuša konstelacijski model v čimvečji meri izkoristiti geometrijsko informacijo. Fergus in sod. (2003), na primer, predstavijo prostorske odnose med značilnicami z vezano verjetnostno porazdelitvijo oblike in izgleda. Pristop sicer omogoča reprezentacijo kategorij z veliko notranjo variabilnostjo oblike in izgleda, vendar kompleksnost in nezanesljivost uporabljenega modela zelo hitro naraščata s številom značilnic. Agarwal in sod. (2004) eksplicitno zakodirajo prostorske odnose med vsemi možnimi pari značilnic, kar povzroči kvadratično naraščanje računske zahtevnosti v odvisnosti od števila značilnic.

Implicitni model oblike upošteva geometrijsko informacijo tako, da zakodira možne položaje vsake posamezne značilnice glede na središča odgovarjajočih objektov, upodobljenih na učnih slikah. Čeprav je takšna predstavitev oblike šibkejša od konstelacijske predstavitve, ISM dopušča precejšnjo variabilnost znotraj kategorije. V Leibe (2004) je navedena uspešna uporaba ISM za detekcijo avtomobilov,

krav, motornih koles in pešcev. V nasprotju z modelom, ki ga zgradijo Fergus in sod. (2003), je model ISM povsem neparametričen. Računska kompleksnost učenja in uporabe ISM linearno narašča s številom značilnic.

Razdelek 3.2 predstavlja učenje in uporabo detekcijskega modela za detekcijo objektov pri konstantni velikosti. Razdelek 3.3 je namenjen razširitvi postopkov iz razdelka 3.2 na detekcijo objektov poljubne velikosti. Razdelek 3.4 natančneje opredeljuje izhod detekcijskega postopka. Ta namreč predstavlja vhod v metodo za vrednotenje rezultatov detekcije (razdelek 3.5), ki je sestavni del detekcijskega modula, ne pa tudi detekcijskega postopka. Razdelek 3.6 sklone poglavje.

3.2 Detekcija objektov konstantne velikosti

V pričujočem razdelku najprej (podrazdelek 3.2.1) na kratko opišemo postopek učenja implicitnega modela oblike za detekcijo objektov pri znani konstantni velikosti. Nato v podrazdelku 3.2.2 prikažemo osnovni postopek detekcije. Podrazdelek 3.2.3 predstavi algoritem, ki pod določenimi pogoji omogoča znaten porast detekcijske točnosti.

Slika 3.1 prikazuje poenostavljeno shemo učnega in testnega postopka za detekcijo objektov pri znani konstantni velikosti.

3.2.1 Učenje detekcijskega modela

Algoritem za učenje detekcijskega modela sprejme množico značilnic $\mathcal{F} = \{F_i\}_i \subseteq \mathcal{Z}$ (\mathcal{Z} podaja univerzalno množico značilnic za dano kategorijo) in množico učnih slik, reprezentiranih z značilnicami iz \mathcal{F} . Vsaka učna slika prikazuje en učni objekt dane vizualne kategorije in je označena z omejevalnim okvirjem objekta. Pri učenju se uporablja le informacija znotraj omejevalnega okvirja. Zaželeno je, če se objekti nahajajo na uniformnem ozadju, da se učni postopek lahko osredotoči zgolj na relevantno informacijo.

Učni postopek zgradi implicitni model oblike za dano kategorijo. Model je določen z množico značilnic \mathcal{F} in s pogojno verjetnostno porazdelitvijo $p(\mathbf{c} | F_i, \mathbf{0})$ za vsako značilnico F_i . Porazdelitev podaja verjetnost, da se središče omejevalnega okvirja objekta nahaja na slikovnih koordinatah \mathbf{c} , če značilnica F_i nastopa na koordinatah $\mathbf{0} = [0, 0]^T$. Zaradi enostavnosti bomo formulacijo „središče omejevalnega okvirja objekta“ nadomestili z ohlapnejšo, a bolj jedrnato formulacijo „središče objekta“.

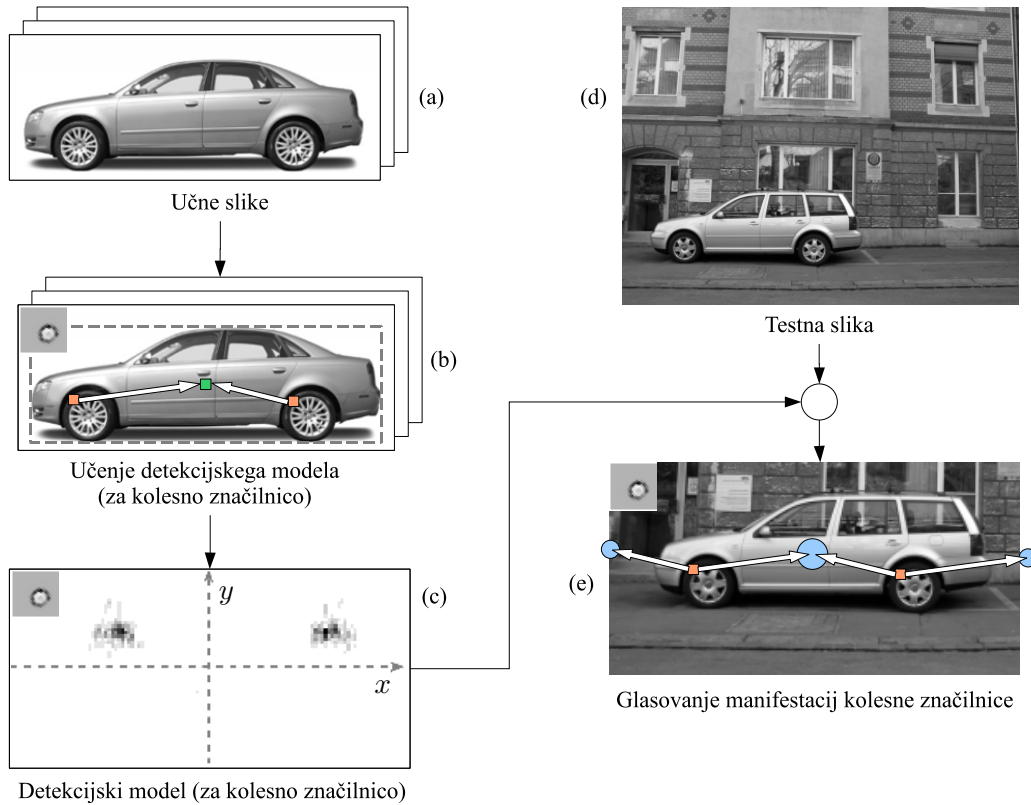
Po temeljni predpostavki (pravzaprav *definiciji*) implicitnega modela oblike za vsak $\mathbf{p} \in \mathbb{R}^2$ velja

$$p(\mathbf{c} + \mathbf{p} | F_i, \mathbf{p}) = p(\mathbf{c} | F_i, \mathbf{0}). \quad (3.1)$$

Pred izvajanjem učnega postopka se učne slike povečajo oziroma pomanjšajo tako, da imajo vsi upodobljeni objekti približno enako velikost. V implementaciji vsaki sliki spremenimo velikost za enak faktor v obeh razsežnostih, tako da objekta ne popačimo. Slike normaliziramo tako, da imajo vsi omejevalni okvirji

bodisi enako širino bodisi enako višino. Če je povprečna višina objektov večja od povprečne širine, normaliziramo na enako višino, sicer pa na enako širino.

Verjetnostno porazdelitev $p(\mathbf{c} | F_i, \mathbf{0})$ je možno oceniti na povsem neparametričen način. Učni postopek za vsako značilnico shrani položaje vseh njenih pojavitev v učnih slikah relativno glede na središča ustrežajočih objektov. Če postopek na sliki, pri kateri se središče objekta nahaja na položaju (slikovnih koordinatah) \mathbf{x} , najde značilnico F_i na položaju \mathbf{y} , potem shrani par $(F_i, \mathbf{x} - \mathbf{y})$. Po sprehodu skozi vse slike se verjetnostna porazdelitev $p(\mathbf{c} | F_i, \mathbf{0})$ aproksimira preprosto kot $n(F_i, \mathbf{c})/n(F_i, \cdot)$, kjer deljenec označuje število shranjenih parov (F_i, \mathbf{c}) , delitelj pa število vseh parov za značilnico F_i . Integracija porazdelitve $p(\mathbf{c} | F_i, \mathbf{0})$ po $\mathbf{c} \in \mathbb{R}^2$



Slika 3.1: Poenostavljena shema učnega (a-c) in testnega (d-e) postopka pri enovelikostni detekciji. **(a)** Učne slike za kategorijo avtomobilov. **(b)** Učenje detekcijskega modela za značilnico, ki se praviloma manifestira na avtomobilskih kolesih. Zeleni kvadrček predstavlja središče omejevalnega okvirja avtomobila. **(c)** Aproksimacija verjetnostne porazdelitve $p(\mathbf{c} | F, \mathbf{0})$ za kolesno značilnico. Temnejša barva pomeni večjo verjetnost. **(d)** Primer testne slike. Širina avtomobila na sliki (d) je dejansko približno enaka širini avtomobila na sliki (a). **(e)** Manifestaciji kolesne značilnice glasujeta za hipoteze o središču avtomobila.

da rezultat 1 za vse F_i , ki se vsaj enkrat manifestirajo v celotni množici učnih slik.

Slika 3.1(c) prikazuje naučeno aproksimacijo porazdelitve $p(\mathbf{c} | F_i, \mathbf{0})$ za značilnico, ki pogosto nastopa na avtomobilskih kolesih. Desna gruča na sliki ustreza pojavitvam te značilnice na kolesih, ki se na učnih upodobitvah nahajajo levo od središč okvirjev avtomobilov, leva gruča pa ustreza kolesom desno od središč okvirjev avtomobilov (slika 3.1b).

Učni postopek je možno enostavno prirediti za primer, če so slike neenolično (verjetnostno) reprezentirane z značilnicami. Pri tej različici se verjetnostne uteži shranijo skupaj s pari ⟨značilnica, relativni položaj⟩, nato pa se uteži upoštevajo tudi pri aproksimaciji verjetnosti.

Poglavitna slabost opisanega postopka je razmeroma velika prostorska zahtevnost, ki je odvisna od števila značilnic in od prostorske ločljivosti oziroma stopnje diskretizacije slikovne ravnine. Postopek izbire značilnic (poglavje 2) tako neposredno vpliva na porabo prostora za ISM v testni fazi. V učni fazi pa je prostor mogoče prihraniti z zmanjšanjem ločljivosti aproksimacije $p(\mathbf{c} | F_i, \mathbf{0})$, pri čemer se tudi neposredni sosedje dane ravninske točke \mathbf{c} pripišejo k aproksimaciji verjetnosti za \mathbf{c} , ali pa s približno izražavo izvirne aproksimacije $p(\mathbf{c} | F_i, \mathbf{0})$ z linearno kombinacijo ravninskih Gaussovih porazdelitev.

3.2.2 Osnovni postopek detekcije

Osnovni detekcijski postopek sprejme sliko, izraženo z značilnicami iz množice $\mathcal{G} = \{G_1, \dots, G_m\} \subseteq \mathcal{F}$, in vrne seznam hipotez, ki napovedujejo položaje morebitnih središč objektov na sliki. Vsaka hipoteza je opremljena s številskim podatkom, imenovanim *podpora* ali *moč*, ki nakazuje stopnjo zaupanja v hipotezo.

Leibe in sod. (2004) so problem detekcije objektov preoblikovali v problem *uteženega glasovanja* za hipoteze, ki napovedujejo morebitna središča objektov v dani testni sliki. Hipoteze prejema glasove s strani manifestacij značilnic na lokalnih strukturah slike. Naj spomnimo, da se pri reprezentaciji slike značilnice prireajo posameznim lokalnim strukturam v sliki. To prirejanje je lahko bodisi enolično – lokalni strukturi se priredi kvečjemu ena značilnica – ali neenolično – lokalni strukturi se lahko priredi poljubno mnogo značilnic, vsaka z določeno verjetnostjo. Lokalna struktura je tako manifestacija kvečjemu ene značilnice (enolična reprezentacija) oziroma poljubnega števila značilnic (neenolična reprezentacija).

Lokalna struktura L , ki se nahaja na slikovnih koordinatah \mathbf{p} , prispeva k hipotezi, ki napoveduje središče objekta na slikovnih koordinatah \mathbf{c} , takšno količino podpore:

$$p(\mathbf{c} | L, \mathbf{p}) = \sum_{i=1}^m p(\mathbf{c} | L, \mathbf{p}, G_i) P(G_i | L, \mathbf{p}). \quad (3.2)$$

Desna stran enačbe ponazarja reprezentacijo lokalne strukture L z značilnicami G_1, \dots, G_m . Ker je pri reprezentaciji lokalna struktura v celoti izražena s temi značilnicami, je v prvem faktorju na desni strani možno odstraniti odvisnost od L . Poleg tega je reprezentacija lokalne strukture z značilnicami (drugi faktor)

neodvisna od nahajališča strukture. Enačba (3.2) se zato poenostavi v

$$p(\mathbf{c} \mid L, \mathbf{p}) = \sum_{i=1}^m p(\mathbf{c} \mid G_i, \mathbf{p}) P(G_i \mid L). \quad (3.3)$$

Faktor $P(G_i \mid L)$ se določi pri reprezentaciji slike z značilnicami, faktor $p(\mathbf{c} \mid G_i, \mathbf{p})$ pa predstavlja verjetnost, da se središče objekta nahaja na absolutnem položaju \mathbf{c} , če značilnica G_i nastopa na absolutnem položaju \mathbf{p} . To verjetnost pa je mogoče pridobiti na podlagi naučenega implicitnega modela oblike. Iz (3.1) namreč sledi

$$p(\mathbf{c} \mid G_i, \mathbf{p}) = p(\mathbf{c} - \mathbf{p} \mid G_i, \mathbf{0}), \quad (3.4)$$

aproksimacija verjetnosti na desni strani pa je rezultat učnega postopka.

Lokalna struktura na položaju \mathbf{p} torej glasuje za vse hipoteze iz množice $\{\mathbf{p} + \mathbf{y} \mid p(\mathbf{y} \mid G_i, \mathbf{0}) > 0\}$, njena podpora posameznim hipotezam pa je podana z enačbo (3.3), če se \mathbf{c} v enačbi nadomesti s $\mathbf{p} + \mathbf{y}$. Slika 3.1e ilustrira glasovanje manifestacij kolesnih značilnic na testni sliki za hipotetična središča omejevalnih okvirjev avtomobilov. Pravilna hipoteza (krog na sredini) prejme največjo podporo, saj zanjo glasujeta obe kolesi. Napačni hipotezi prejmeta manjšo podporo, saj za levo hipotezo glasuje le levo kolo, za desno hipotezo pa zgolj desno kolo.

Podpora, ki jo hipotezi \mathbf{c} namenijo vse lokalne strukture v sliki skupaj, znaša

$$p(\mathbf{c}) = \sum_k p(\mathbf{c} \mid L_k, \mathbf{p}_k) = \sum_k \sum_{i=1}^m p(\mathbf{c} - \mathbf{p}_k \mid G_i, \mathbf{0}) P(G_i \mid L_k), \quad (3.5)$$

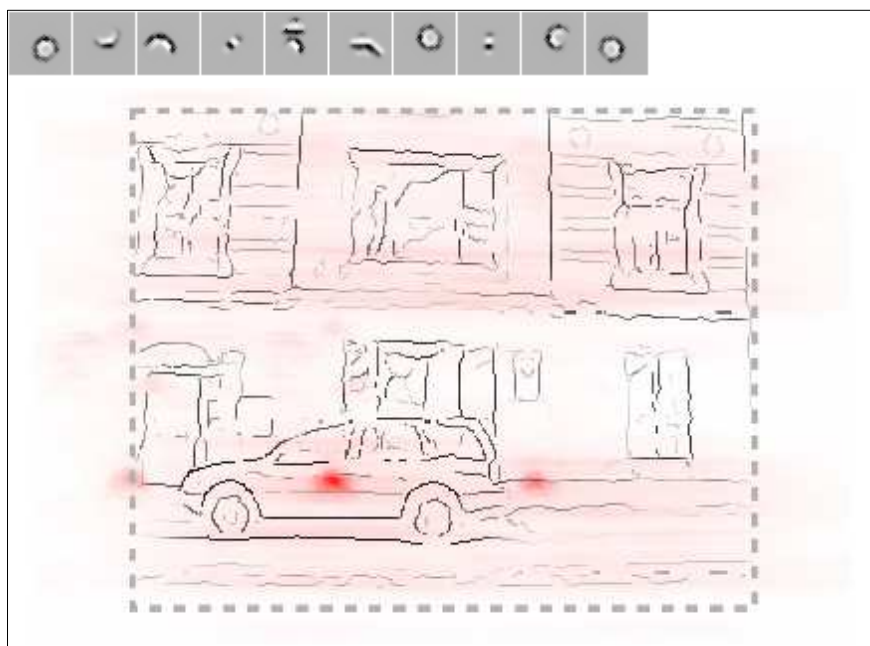
če struktura L_k nastopa na položaju \mathbf{p}_k . Množica vrednosti $p(\mathbf{c})$ nad celotno domeno spremenljivke \mathbf{c} (t.j. \mathbb{R}^2) določa *hipotezni prostor*. Na sliki 3.2 je prikazan hipotezni prostor, pridobljen z izvajanjem detekcijskega postopka na sliki 3.1d.

Zaradi variabilnosti izgleda objektov v kategoriji ni mogoče pričakovati, da se bodo vsi glasovi za nek objekt nakopičili v eni točki v \mathbb{R}^2 . Detekcijski postopek zato vrne množico *stabilnih* detekcijskih hipotez, ki jo je možno pridobiti z uporabo postopka *mean-shift* (Comaniciu in Meer, 1999) nad hipoteznim prostorom ali pa preprosto z iskanjem lokalnih maksimumov v hipoteznem prostoru, zglajenem z Gaussovim filtrom. Izkaže se, da slednja rešitev v praksi povsem zadošča. V implementaciji smo uporabili Gaussovo jedro velikosti 11×11 z vrednostjo $\sigma = 2$.

Ker tako *mean-shift* kot Gaussovo filtriranje v osnovi zgladita prostor z neko jedrno funkcijo, je moč dane končne hipoteze (stabilnega lokalnega maksimuma) mogoče izraziti kot

$$s(\mathbf{c}) = \sum_{\mathbf{c}_j \in W(\mathbf{c})} g(\mathbf{c}_j - \mathbf{c}) p(\mathbf{c}_j), \quad (3.6)$$

pri čemer $W(\mathbf{c})$ označuje okno s središčem v točki \mathbf{c} , ki ga določa jedrna funkcija, faktor $g(\mathbf{c}_j - \mathbf{c})$ pa podaja vrednost jedrne funkcije v točki \mathbf{c}_j v okolici \mathbf{c} . Za vse točke $\mathbf{y} \in \mathbb{R}^2$, ki niso v množici končnih hipotez, velja $s(\mathbf{y}) = 0$.



Slika 3.2: Hipotezni prostor (pred glajenjem), dobljen z izvajanjem osnovnega detekcijskega postopka na testni sliki 3.1d. Postopek uporablja le prikazane značilnice. Zaradi lažje predstave je hipotezni prostor superponiran na reprezentacijo testne slike, pridobljeno z detektorjem robov tipa Berkeley (Martin in sod., 2004). Črtkani okvir predstavlja meje slike. Najmočnejša hipoteza se nahaja približno v središču avtomobila in je zato tudi pravilna. Preostali del hipoteznega prostora večinoma izvira iz ozadja. Izjemi sta nekoliko močnejši gruči neposredno levo in neposredno desno od avtomobila, ki izvirata iz levega oziroma desnega kolesa.

3.2.3 Dušenje sekundarnih hipotez

Osnovni detekcijski algoritem sicer pogosto najde močno hipotezo, ki se ujema z dejanskim središčem objekta v sliki, a vrne tudi mnogo močnih napačnih hipotez. Detekcijska točnost preko celotne testne množice je posledično v mnogih primerih precej nizka.

Z algoritmom za dušenje sekundarnih hipotez, ki je v nadaljnjem besedilu imenovan „algoritem MDL“, je moč izboljšati detekcijsko točnost, če sta izpolnjena naslednja pogoja:

1. Najmočnejša detekcijska hipoteza je pogosto pravilna. Če torej testna slika vsebuje vsaj en objekt naučene kategorije, potem bo najmočnejša hipoteza najverjetneje pravilno napovedala položaj enega od objektov.
2. Detekcijski postopek vrne veliko število močnih napačnih hipotez, ki morda

niso najmočnejše znotraj slik, iz katerih izvirajo, so pa močnejše od mnogih pravih hipotez v celotni slikovni množici.

Glasovalna shema omogoča, da ista lokalna struktura podpre več kot eno hipotezo v dani sliki. Algoritem MDL pa temelji na predpostavki, da je vsako lokalno strukturo smiselno povezati s kvečjemu eno hipotezo. Hipoteza napoveduje položaj nekega objekta dane kategorije v sliki, vzrok njenega nastanka pa so deli objekta ali ozadja, ki kot lokalne strukture oz. manifestacije značilnic glasujejo zanjo. Nek del, denimo kolo avtomobila, sicer lahko *glasuje* za več kot eno hipotezo, torej za več možnih položajev avtomobilov, vendar pa po predpostavki algoritma MDL *pripada* kvečjemu enemu dejanskemu avtomobilu na sliki. Za uresničitev predpostavke enolične pripadnosti mora algoritem MDL izhodiščno stanje, v katerem se lahko zgodi, da ista lokalna struktura podpira več hipotez, pretvoriti v stanje, kjer vsaka lokalna struktura podpira največ eno hipotezo. To pretvorbo izvrši na požrešen iterativen način.

Pred nadaljnjo obravnavo je koristno definirati pojem *vzratne projekcije* (angl. backprojection po Leibeju in sod.) dane hipoteze \mathbf{c} :

$$\mathcal{B}(\mathbf{c}) = \{(L_k, \mathbf{p}_k) \mid \sum_{\mathbf{c}_j \in W(\mathbf{c})} g(\mathbf{c}_j - \mathbf{c}) \sum_k p(\mathbf{c}_j \mid L_k, \mathbf{p}_k) > 0\}. \quad (3.7)$$

Gre torej za množico lokalnih struktur, ki podpirajo dano hipotezo. Po zaključku izvajanja algoritma MDL so si vse množice $\mathcal{B}(\mathbf{c})$ medsebojno tuje, saj se vsaka lokalna struktura priredi le eni hipotezi.

Opredelimo še pojem *prispevka* (angl. contribution) množice lokalnih struktur $\mathcal{L} = \{(L_k, \mathbf{p}_k)\}_k$ k dani hipotezi:

$$c(\mathcal{L}, \mathbf{c}) = \sum_{\mathbf{c}_j \in W(\mathbf{c})} g(\mathbf{c}_j - \mathbf{c}) \sum_{(L, \mathbf{p}) \in \mathcal{L}} p(\mathbf{c}_j \mid L, \mathbf{p}). \quad (3.8)$$

Velja torej $s(\mathbf{c}) = c(\mathcal{B}(\mathbf{c}), \mathbf{c})$.

Algoritem MDL obravnava vsako sliko posebej. V prvi iteraciji se osredotoči na najmočnejšo hipotezo v sliki, \mathbf{c}^* , in vse lokalne strukture iz množice $\mathcal{B}(\mathbf{c}^*)$ priredi *izključno* hipotezi \mathbf{c}^* . S tem moč najmočnejše hipoteze ostane nespremenjena, vse ostale hipoteze pa izgubijo celotno količino podpore, ki so jim jo dodelile lokalne strukture iz množice $\mathcal{B}(\mathbf{c}^*)$:

$$\begin{aligned} \mathcal{B}'(\mathbf{c}) &= \mathcal{B}(\mathbf{c}) \setminus \mathcal{B}(\mathbf{c}^*), \\ s'(\mathbf{c}) &= c(\mathcal{B}'(\mathbf{c}), \mathbf{c}) \end{aligned} \quad (3.9)$$

za vse hipoteze $\mathbf{c} \neq \mathbf{c}^*$. Relativna izguba moči dane hipoteze \mathbf{c} je torej odvisna od količine podpore, ki je bila hipoteza deležna s strani lokalnih struktur v množici $\mathcal{B}(\mathbf{c}) \cap \mathcal{B}(\mathbf{c}^*)$, v primerjavi s podporo s strani struktur v izvorni množici $\mathcal{B}(\mathbf{c})$.

V naslednji iteraciji algoritem hipoteze \mathbf{c}^* in lokalnih struktur iz $\mathcal{B}(\mathbf{c}^*)$ ne obravnava več. Osredotoči se na hipotezo, ki je izmed preostalih hipotez trenutno

najmočnejša, in ji požrešno priredi vse strukture iz njene vzvratne projekcije, s čimer oslabi druge hipoteze na način, kot ga določata enačbi (3.9). Postopek se tako ponovi. Algoritem MDL se ustavi po obravnavi vseh hipotez oziroma tedaj, ko so vse lokalne strukture že prirejene kvečjemu eni hipotezi. Algoritem MDL je formalneje prikazan kot algoritem 3.1.

Algoritem 3.1 Algoritem MDL pri enovelikostni detekciji.

Vhod:

- (a) Učna slika, reprezentirana z značilnicami iz \mathcal{G} .
- (b) Množica detekcijskih hipotez $\{\mathbf{c}\}_{\mathbf{c}}$.

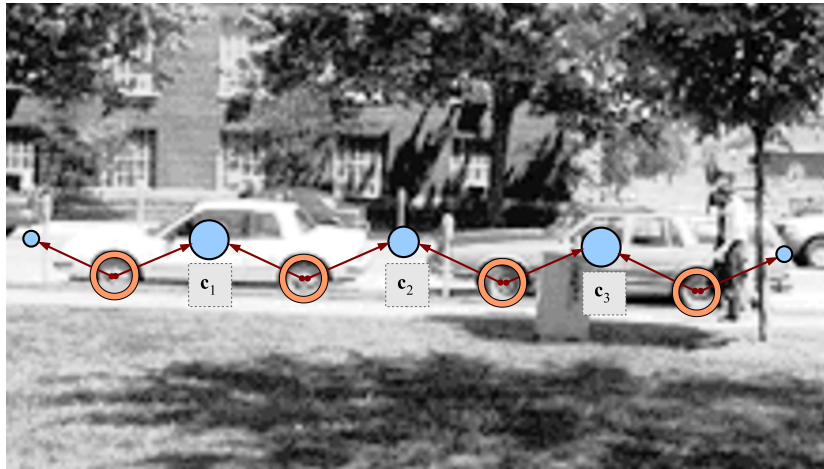
Izhod: Množica istih detekcijskih hipotez s posodobljenimi vrednostmi $s(\mathbf{c})$.

Postopek:

- 1: Določi $\mathcal{B}(\mathbf{c})$ za vsako hipotezo \mathbf{c} . Po definiciji velja $s(\mathbf{c}) = c(\mathcal{B}(\mathbf{c}), \mathbf{c})$.
 - 2: $\mathcal{H} := \{\mathbf{c}\}_{\mathbf{c}}$;
 - 3: **dokler** $\mathcal{H} \neq \emptyset$ **ponavljaj**
 - 4: $\mathbf{c}^* := \arg \max_{\mathbf{c} \in \mathcal{H}} s(\mathbf{c})$;
 - 5: **za vse** $\mathbf{c} \neq \mathbf{c}^*$ **ponavljaj**
 - 6: $\mathcal{B}(\mathbf{c}) := \mathcal{B}(\mathbf{c}) \setminus \mathcal{B}(\mathbf{c}^*)$;
 - 7: $s(\mathbf{c}) := c(\mathcal{B}(\mathbf{c}), \mathbf{c})$
 - 8: **konec** **{za}**;
 - 9: $\mathcal{H} := \mathcal{H} \setminus \{\mathbf{c}^*\}$
 - 10: **konec** **{dokler}**;
 - 11: **vrni** množico hipotez $\{\mathbf{c}\}_{\mathbf{c}}$.
-

Leibe in sod. so uporabo algoritma MDL pri vizualni detekciji motivirali z zanimivim primerom. Značilnice, ki robustno zaznavajo avtomobilska kolesa, se izkažejo kot zelo koristne za detekcijo avtomobilov v pogledu od strani. Ker pa se sprednje kolo po izgledu običajno nič ne razlikuje od zadnjega, ima naučena porazdelitev $p(\mathbf{c} | G_i)$ za kolesne značilnice dva izrazita maksimuma (slika 3.1c). Zato v testni fazi zaznano kolo praviloma podpre dve stabilni hipotezi, eno pravilno in eno napačno (slika 3.1e). Če se na sliki nahajata dva avtomobila v razmestitvi, ki jo prikazuje slika 3.3, potem v prostoru med obema avtomobiloma nastane močna napačna hipoteza, ki je na sliki označena kot \mathbf{c}_2 . Brez uporabe algoritma MDL bi hipoteza \mathbf{c}_2 gotovo negativno vplivala na detekcijsko točnost. Algoritem MDL pa pri pogojih, ki so domala vedno izpolnjeni, hipotezo \mathbf{c}_2 (skoraj) povsem zaduši, hipotezi \mathbf{c}_1 in \mathbf{c}_3 pa pusti nedotaknjeni.

Če je pred pričetkom izvajanja algoritma MDL najmočnejša hipoteza \mathbf{c}_1 , jo bo algoritem izbral v prvi iteraciji in s tem močno oslabil hipotezo \mathbf{c}_2 , saj sta \mathbf{c}_2 in \mathbf{c}_1 prejeli precej podpore od skupnih lokalnih struktur. Tudi če je hipoteza \mathbf{c}_2 bila v izhodišču močnejša od \mathbf{c}_3 , bo po dušilnem koraku hipoteza \mathbf{c}_3 , ki z izbiro \mathbf{c}_1 ni izgubila ničesar, skoraj gotovo imela večjo podporo kot \mathbf{c}_2 . Tako bo algoritem v drugi iteraciji izbral hipotezo \mathbf{c}_3 , s čimer bo \mathbf{c}_2 izgubila še več podpore in bo tako postala zanemarljivo šibka. Podoben razmislek velja za primer, če je pred pričetkom algoritma najmočnejša hipoteza \mathbf{c}_3 namesto \mathbf{c}_1 . Razmere bi bile



Slika 3.3: Primer močne napačne hipoteze (c_2), ki jo algoritem MDL uspešno zaduši.

drugačne, če bi hipoteza c_2 bila v izhodišču močnejša od c_1 in c_3 . Vendar se to skoraj nikoli ne zgodi, saj c_1 in c_3 črpata podporo še iz drugih delov obeh avtomobilov, ne zgolj iz koles, ki so poglavitni vir podpore za c_2 .

Predstavljeni algoritem so razvili Leonardis in sod. (1995) kot pristop k iskanju modela, ki doseže najboljši kompromis med lastno kompleksnostjo in natančnostjo opisa dane množice podatkov. Gre za kriterij po načelu najmanjše dolžine opisa (Minimum Description Length, MDL). Ker je problem izbire optimalnega modela zelo splošen, je bil algoritem MDL uporabljen v različnih domenah (Leonardis in sod., 2002; Bischof in sod., 1999). Leibe in sod. (2004) so ga uspešno prenesli na domeno detekcije objektov.

3.3 Detekcija objektov poljubne velikosti

Opisani detekcijski postopek je ne glede na vrsto uporabljenih značilnic mogoče premočrtno razširiti za uporabo v primeru, ko testna slikovna množica vsebuje podobe objektov ciljne kategorije v poljubni velikosti. V podrazdelku 3.3.1 predstavljamo precej standarden načrt za detekcijo pri poljubni velikosti. Ta načrt smo uresničili pri uporabi značilnic avtorjev Fidler in sod. (2006).

Sestavi značilnic, definirani na velikostno invarianten način, pa omogočajo tudi drugačne strategije za razširitev detekcije na večvelikostno domeno. Razširitev, ki so jo predlagali Leibe in sod., mi pa smo se je poslužili pri uporabi značilnic avtorjev Ferrari in sod. (2006a), je opisana v podrazdelku 3.3.2.

3.3.1 Splošna shema večvelikostne detekcije

Shemo, predstavljeno v tem podrazdelku, je možno uporabiti v kombinaciji s poljubno vrsto značilnic. Podoben načrt za razširitev izhodiščnega enovelikostnega detekcijskega algoritma na večvelikostno domeno je izbralo že več avtorjev (Agarwal in sod., 2004; Viola in Jones, 2004). Vložitev (angl. embedding) slike ali splošnega signala v velikostni prostor (angl. scale space), kar predstavlja poglaviti sestavni del sheme, sta teoretsko obdelala Koenderink (1984) in Lindeberg (1994).

3.3.1.1 Učni postopek

Učni algoritem je povsem tak kot v primeru enovelikostne detekcije. Nespremenjena ostane tudi predpostavka, da so vsi objekti v učni slikovni množici upodobljeni v približno enaki velikosti in da so slike opremljene s podatki o omejevalnih okvirjih objektov.

3.3.1.2 Osnovni detekcijski algoritem (brez postopka MDL)

Detekcijski postopek za vsako testno sliko zgradi piramido v velikostnem prostoru (angl. scale space pyramid) (Crowley, 1981). Piramida je razdeljena na *oktave*, vsaka oktava pa zajema eno ali več *plastí*. Število plasti na oktavo je konstantno in določeno vnaprej. Vsaka plast vsebuje povečano oz. pomanjšano kopijo izvirne slike. Naj 0 označuje plast, ki vsebuje izvirno sliko, R pa število plasti na oktavo. Potem plast z oznako $r \in \mathbb{Z}$ vsebuje sliko z velikostjo enako $2^{r/R}$ -kratniku velikosti izvirne slike. Prehod za eno oktavo po piramidi navzdol tako pomeni 2-kratno povečanje slike v dani oktavi, prehod za eno oktavo po piramidi navzgor pa 2-kratno zmanjšanje te slike. Primer piramide v velikostnem prostoru je prikazan na sliki 3.4.

Pred spreminjanjem ločljivosti je izvirno sliko koristno zgladiti z nizkoprepustnim filtrom, da se zmanjša učinek „presluha“ različnih prostorskih frekvenc v sliki (angl. aliasing). V implementaciji smo uporabili Matlabovo funkcijo `imresize`, ki sliko najprej zgladi z Gaussovim filtrom primerne velikosti, nato pa jo z (bikubično) interpolacijo poveča ali zmanjša na ciljno velikost.

Osnovni detekcijski algoritem se nato požene na vsaki plasti posebej. Ker učni algoritem proizvede reprezentacijo, vezano na učno velikost, je pravilno hipotezo za dani objekt mogoče pričakovati kvečjemu na plasti, na kateri je objekt upodobljen v velikosti, ki je približno enaka učni velikosti. Uspeh detekcije je tako odvisen tudi od ločljivosti v velikostnem prostoru, t.j. od števila plasti na oktavo. Večja vrednost R pomeni večjo možnost pravilne detekcije, a tudi večjo računsko zahtevnost. V implementaciji je število plasti na oktavo v vseh primerih enako 4. Velikostni faktor med sosednjima plastema potemtakem znaša $\sqrt[4]{2} \approx 1,19$.

Število plasti v piramidi mora biti dovolj veliko, da pokrije celoten pričakovani razpon velikosti testnih objektov. Seveda je treba vsako testno sliko obravnavati na celotni piramidi, saj iz velikosti slike ni mogoče sklepati na velikost objekta.

Če pa velja predpostavka, da velikost objekta nikoli ne presega velikosti slike, potem je smiselno izločiti plasti, na katerih slika postane manjša od najmanjše pričakovane velikosti testnega objekta.

3.3.1.3 Prilagoditev postopka MDL

Osnovni detekcijski algoritem tvori hipoteze za vsako plast piramide posebej. Zato bi tudi algoritem MDL bilo možno pognati na vsaki plasti ločeno in tako udušiti sekundarne hipoteze pri posameznih velikostih. Vendar bi na tak način najmočnejše hipoteze posameznih plasti ostale nedotaknjene, zaradi česar postopek MDL ne bi nujno dosegel želenega učinka.

Plasti niso neodvisne med seboj, saj vsaka vsebuje povečano ali pomanjšano kopijo iste izvirne slike. Zaradi tega se isti del objekta (npr. kolo avtomobila) odraža na več plasteh, čeprav je pri sestavu značilnic, občutljivem na velikostne spremembe, reprezentacija istega dela objekta lahko različna od plasti do plasti.

Različica algoritma MDL, ki upošteva opisano odvisnost med plastmi, v vsaki iteraciji izbere trenutno najmočnejšo hipotezo v *celotni* piramidi. Denimo, da gre za hipotezo \mathbf{c}^* na plasti r^* . Podobno kot v enovelikostnem primeru se zatem hipotezi \mathbf{c}^* požrešno priredijo vse lokalne strukture na plasti r^* , ki so zanjo glasovale. Posledično se zmanjša podpora konkurenčnim hipotezam na plasti r^* . Kot je navedeno v prejšnjem odstavku, požrešno prirejene lokalne strukture na plasti r^* ustrezajo določenim delom izvirne slike, ki se v splošnem odražajo tudi na drugih plasteh, ne zgolj na plasti r^* . Če algoritem MDL izbrani hipotezi \mathbf{c}^* priredi npr. avtomobilsko kolo na plasti r^* , mora iz nadaljnje obravnave odstraniti kolesu ustrezajoče lokalne strukture na vseh plasteh, saj se s požrešnim prirejanjem kolo enolično in dokončno poveže s hipotezo \mathbf{c}^* .

Če značilnice niso velikostno invariantne, potem lokalnih struktur na različnih plasteh ni mogoče neposredno primerjati. Zato je težko ali nemogoče natančno določiti, katere lokalne strukture na poljubni plasti r ustrezajo dani strukturi na plasti r^* . Zaradi tega se algoritem MDL posluži naslednje hevristike: Za vsako lokalno strukturo L na plasti r^* , ki je bila prirejena hipotezi \mathbf{c}^* , poišče na vsaki od preostalih plasti vse lokalne strukture, ki se nahajajo na približno enakem relativnem položaju znotraj ustreznih slik kot struktura L znotraj slike na plasti r^* . Če se struktura L na sliki plasti r^* nahaja na položaju (x^*, y^*) , ji na plasti r po tej hevristiki ustrezajo strukture na položajih

$$\mathcal{P}(r, r^*, (x^*, y^*)) = \mathcal{N}(2^{(r-r^*)/R}(x^*, y^*), \lceil 2^{(r-r^*)/R} \rceil), \quad (3.10)$$

kjer $\mathcal{N}((p, q), \varepsilon)$ označuje krožno okolico okrog točke (p, q) s polmerom ε . Vsaka tako dobljena lokalna struktura se požrešno priredi hipotezi \mathbf{c}^* , s čimer se moč hipoteze \mathbf{c}^* ne spremeni, zmanjša pa se moč hipotez na plasteh, iz katerih izvirajo posamezne dobljene strukture. Ni nujno, da vse strukture iz množice $\mathcal{P}(r, r^*, (x^*, y^*))$ dejansko ustrezajo strukturi (x^*, y^*) na plasti r^* , vendar se uporabljena hevristika v praksi dobro obnese.

3.3.2 Večvelikostna detekcija pri velikostno invariantnih značilnicah

Sestavi velikostno invariantnih značilnic, kot so npr. značilnice SIFT (Lowe, 2004) ali značilnice avtorjev Ferrari in sod. (2006a), omogočajo tudi drugačne detekcijske strategije. V idealnem primeru je vsako značilnico iz tovrstnega sestava mogoče ponovljivo zaznati pri poljubni velikosti. Ob predpostavki takšnih razmer zadošča, če tako učni kot testni algoritem obravnavata slike le pri eni velikosti, ki je lahko poljubna in od slike do slike drugačna. V nadaljevanju predpostavljamo popolno velikostno invariantnost sestava značilnic, ob koncu razdelka pa bomo navedli, kako je mogoče postopati v manj idealnih razmerah.

Shemo, opisano v besedilu do zaključka tega razdelka, so predlagali Leibe in sod. (2004). Slika 3.4 ilustrira učenje in detekcijo na podlagi velikostno invariantnih značilnic.

3.3.2.1 Učni postopek

Učni postopek sprejme množico učnih slik z znanimi omejevalnimi okvirji. Objekti so lahko upodobljeni v poljubni velikosti. Za razliko od enovelikostnega učnega postopka se v tem primeru poleg položajev posameznih pojavitev značilnic obravnavajo tudi njihove *relativne velikosti*. V predpostavljenih idealnih razmerah je relativna velikost vsake pridobljene značilnice premosorazmerna velikosti slike. Relativna velikost značilnice je podana kot večkratnik neke izhodiščne velikosti, ki je določena že v fazi pridobivanja značilnic.

Denimo, da se središče objekta na dani učni sliki nahaja na slikovnih koordinatah \mathbf{x} , relativna velikost objekta pa znaša s . (Relativna velikost objekta je določena kot večkratnik neke fiksne izhodiščne učne velikosti.) Potem učni algoritem ob pojavitvi značilnice F_i na položaju \mathbf{y} in v relativni velikosti t shrani trojico

$$(F_i, \frac{\mathbf{x} - \mathbf{y}}{t}, \frac{s}{t}). \quad (3.11)$$

Po pregledu vseh slik se za vsako značilnico F_i oceni porazdelitev $p(\mathbf{c}, v | F_i, \mathbf{0}, 1)$, ki ponazarja vezano verjetnost položaja objektovega središča (\mathbf{c}) in relativne objektove velikosti (v) pri pogoju, da značilnica F_i nastopa na slikovnih koordinatah $\mathbf{0} = [0, 0]^T$ in v relativni velikosti 1. Po predpostavki razširjenega implicitnega modela oblike za vsak $t > 0$ in $\mathbf{p} \in \mathbb{R}^2$ velja

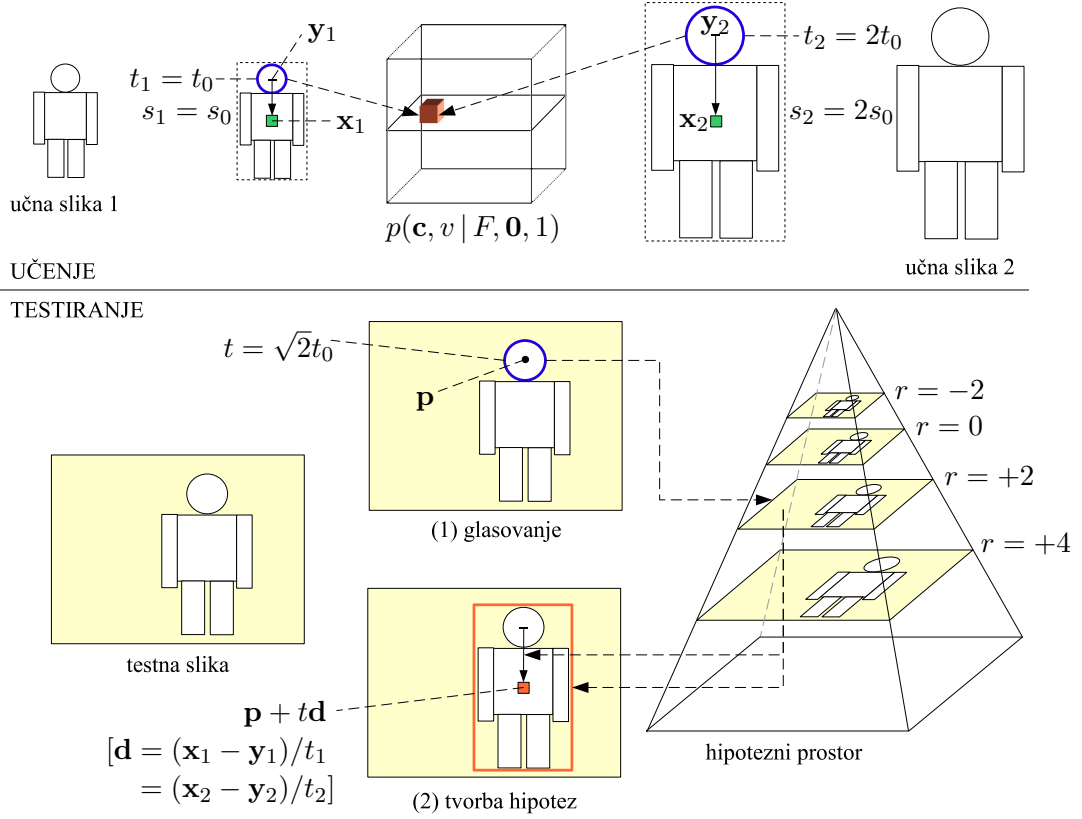
$$p(t\mathbf{c} + \mathbf{p}, tv | F_i, \mathbf{p}, t) = p(\mathbf{c}, v | F_i, \mathbf{0}, 1). \quad (3.12)$$

Porazdelitev $p(\mathbf{c}, v | F_i, \mathbf{0}, 1)$ se aproksimira na podoben način kot porazdelitev $p(\mathbf{c} | F_i, \mathbf{0})$ pri enovelikostnem učnem postopku.

3.3.2.2 Detekcijski postopek

Ob predpostavki popolne velikostne invariantnosti značilnic zadostuje, če detekcijski postopek obravnava testno sliko zgolj v njeni izvorni velikosti. Izgradnja

velikostne piramide ni potrebna. Detekcijski algoritem za dano testno sliko tvori



Slika 3.4: Poenostavljen in idealiziran prikaz večvelikostne detekcije z velikostno invariantnimi značilnicami. Denimo, da detekcija temelji le na značilnici, ki se odziva na krožne strukture poljubne velikosti. Velikost (npr. višina) prve učne slike naj znaša $s_1 = s_0$, velikost druge učne slike pa naj bo $s_2 = 2s_0$. Središči objektov naj se nahajata v točkah \mathbf{x}_1 oz. \mathbf{x}_2 . Značilnica naj se na prvi sliki manifestira na položaju \mathbf{y}_1 v velikosti $t_1 = t_0$, na drugi sliki pa na položaju \mathbf{y}_2 v velikosti $t_2 = 2t_0$. Ker velja $(\mathbf{x}_1 - \mathbf{y}_1)/t_1 = (\mathbf{x}_2 - \mathbf{y}_2)/t_2$ in $s_1/t_1 = s_2/t_2$, učni algoritem za obe manifestaciji shrani identično trojico (3.11). Denimo, da se na testni sliki značilnica manifestira na položaju \mathbf{p} v velikosti $t = \sqrt{2}t_0$. V skladu s (3.15) manifestacija glasuje za položaj $\mathbf{p} + t\mathbf{d}$, kjer je $\mathbf{d} = (\mathbf{x}_1 - \mathbf{y}_1)/t_1 = (\mathbf{x}_2 - \mathbf{y}_2)/t_2$, in za velikost objekta $t(s_1/t_1) = t(s_2/t_2) = \sqrt{2}s_0$, ki ustreza piramidni plasti $r = +2$ oziroma oranžnemu omejevalnemu okvirju. OPOMBA: Piramida na sliki služi zgolj kot ilustracija hipoteznega prostora. Ta dejansko vsebuje glasove za hipoteze, ne povečanih in pomanjšanih kopij testne slike, kot bi bilo mogoče sklepati na podlagi ilustracije. Testna slika se v tem primeru obravnava le pri njeni izvirni velikosti in ne v okviru velikostne piramide.

hipoteze o središčih objektov in njihovih velikostih, relativnih glede na izhodiščno velikost v učni fazi. Medtem ko so v različici iz podrazdelka 3.3.1 hipotetične velikosti objektov določene posredno, preko podatka o plasteh posameznih lokacijskih hipotez, je v različici detekcijskega algoritma po Leibeju in sod. relativna velikost neločljivi del vsake hipoteze. Prostor hipotez je tako trirazsežen: dve razsežnosti za položaj in ena za velikost.

Po analogiji z enačbo (3.3) prispeva lokalna struktura L na položaju \mathbf{p} k hipotezi (\mathbf{c}, v) sledečo količino podpore:

$$p(\mathbf{c}, v | L, \mathbf{p}) = \sum_{i=1}^m p(\mathbf{c}, v | G_i, \mathbf{p}, t_i) P(G_i | L), \quad (3.13)$$

kjer (\mathbf{c}, v) označuje skupno hipotezo o položaju središča objekta in relativni velikosti objekta, t_i pa predstavlja relativno velikost značilnice G_i , ki sodeluje pri reprezentaciji lokalne strukture L . Prvi faktor na desni strani enačbe podaja verjetnost hipoteze (\mathbf{c}, v) , če se na položaju \mathbf{p} nahaja značilnica G_i v relativni velikosti t_i . Ta verjetnost se po predpostavki (3.12) preoblikuje v porazdelitveno gostoto, ki jo je aproksimiral učni algoritem:

$$p(\mathbf{c}, v | G_i, \mathbf{p}, t_i) = p\left(\frac{\mathbf{c} - \mathbf{p}}{t_i}, \frac{v}{t_i} | G_i, \mathbf{0}, 1\right). \quad (3.14)$$

Lokalna struktura L na položaju \mathbf{p} glasuje za vse hipoteze iz množice

$$\{(\mathbf{p} + t_i \mathbf{d}, t_i w) \mid (p(\mathbf{d}, w | G_i, \mathbf{0}, 1) > 0) \wedge (L \text{ je manifestacija } G_i \text{ v velikosti } t_i)\}, \quad (3.15)$$

kar neposredno sledi iz enačb (3.12) in (3.13).

Prispevek vseh lokalnih struktur k dani hipotezi se izračuna po formuli

$$p(\mathbf{c}, v) = \sum_k p(\mathbf{c}, v | L_k, \mathbf{p}_k). \quad (3.16)$$

Verjetnosti $p(\mathbf{c}, v)$ nad \mathbb{R}^3 tvorijo hipotezni prostor.

Zaradi variabilnosti izgleda objektov v kategoriji ni mogoče pričakovati, da se bodo vsi glasovi za kombinacijo položaja in velikosti nekega objekta nakopičili v isti trirazsežni točki. Leibe in sod. poiščejo stabilne detekcijske hipoteze z uporabo neparametričnega postopka za ocenjevanje verjetnostne gostote, ki je osnovan na prilagojeni različici algoritma *mean-shift*. V naši implementaciji smo se poslužili preprostejšega pristopa, ki ga opisujemo v nadaljevanju.

Hipotezni prostor najprej diskretiziramo v vseh treh razsežnostih. V implementaciji smo razsežnosti x in y , ki ustrezata slikovnim koordinatam, diskretizirali tako, da ena dvorazsežna celica ustreza kvadratu s stranico, dolgo 2 slikovna elementa. Tretjo, torej velikostno razsežnost smo diskretizirali v smislu piramidalne reprezentacije. Velikostno os smo tako razdelili na plasti, pri čemer smo razmik

med sosednjima plastema nastavili na četrtno oktave ($R = 4$). Plast 0 ustreza izhodiščni velikosti, določeni v učnem algoritmu. Plast $r \in \mathbb{Z}$ ustreza relativni velikosti $v = 2^{r/R}$ glede na izhodiščno velikost. (Zaradi diskretizacije se na plasti r zberejo glasovi za vse relativne velikosti, ki so bližje vrednosti $2^{r/R}$ kot vrednostim $2^{(r-1)/R}$ in $2^{(r+1)/R}$.)

V fazi glasovanja se prispevki lokalnih struktur zberejo v tako diskretiziranem prostoru. Izhajamo iz predpostavke, da se stabilne detekcijske hipoteze odražajo kot lokalni maksimumi v zglajenem glasovalnem prostoru. Po tej predpostavki je vsaka hipoteza skupek vseh glasov v njeni trirazsežni okolici.

Vsako plast hipoteznega prostora zgladimo s trirazsežnim Gaussovim jedrom. Podpore hipotezam na dani plasti tako prispevajo tudi glasovi s sosednjih plasti. Stopnja glajenja v velikostni osi je konstantna za celoten hipotezni prostor, v oseh x in y pa je odvisna od plasti. Pri plasteh, na katerih se zbirajo glasovi za večje velikosti objektov, je stopnja glajenja v oseh x in y večja, saj je na takih plasteh večja tudi položajna variabilnost. To odločitev je moč utemeljiti tudi z dejstvom, da je premik za eno celico v smeri x ali y na plasti za manjšo velikost relativno bolj opazen kot enak premik na plasti za večjo velikost. Sledi definicija Gaussovega filtra, ki smo ga uporabili za glajenje posameznih plasti:

$$g(x, y, r) = \underbrace{\exp\left(-\frac{x^2}{2\sigma_x(r)^2}\right)}_{g_x(x, y, r)} * \underbrace{\exp\left(-\frac{y^2}{2\sigma_y(r)^2}\right)}_{g_y(x, y, r)} * \underbrace{\exp\left(-\frac{r^2}{2\sigma_r^2}\right)}_{g_r(x, y, r)}, \quad (3.17)$$

pri čemer $*$ označuje trirazsežno diskretno konvolucijo:

$$h_3(x, y, r) = \sum_{x'=-\infty}^{\infty} \sum_{y'=-\infty}^{\infty} \sum_{r'=-\infty}^{\infty} h_1(x', y', r') h_2(x - x', y - y', r - r'), \quad (3.18)$$

če velja $h_3 = h_1 * h_2$. Dejansko konvolucije ni treba računati, saj je zaradi separabilnosti možno aplikacijo trirazsežnega Gaussovega filtra na hipotezni prostor nadomestiti z aplikacijami ustreznih enorazsežnih komponent, ki so v enačbi (3.17) označene z g_x , g_y in g_r . Vsaka plast hipoteznega prostora se tako najprej filtrira z g_x , potem z g_y , nazadnje pa še z g_r . Vrstni red ni pomemben, pri filtriranju z g_r pa poleg trenutne plasti sodelujejo tudi druge plasti. V implementaciji smo konstanten parameter σ_r postavili na 1, parametra $\sigma_x(r)$ in $\sigma_y(r)$ pa na $(3\sqrt{2^{r/R}})$.

Stabilne hipoteze se nato določijo kot trirazsežni lokalni maksimumi v zglajenem hipoteznem prostoru. Točka (x, y, r) je trirazsežni lokalni maksimum, če je vrednost v njej večja ali enaka kot vrednosti v vseh 26 sosednjih točkah. Z upoštevanjem glajenja je torej moč dane stabilne hipoteze enaka

$$s((x, y), 2^{r/R}) = \sum_{(x_j, y_j, r_j) \in W(x, y, r)} g(x_j - x, y_j - y, r_j - r) p((x_j, y_j), 2^{r_j/R}), \quad (3.19)$$

kjer $W(x, y, r)$ označuje kvader vpliva Gaussove funkcije g s središčem v točki (x, y, r) . Dolžina stranice kvadra v dani razsežnosti je tipično 5-kratnik parametra glajenja σ v tej razsežnosti.

3.3.2.3 Postopek MDL

Opisano detekcijsko shemo je mogoče nadgraditi s postopkom za dušenje sekundarnih hipotez na povsem enak način kot pri osnovnem detekcijskem postopku (podrazdelek 3.2.3). Edina razlika je v predstavitvi hipotez, ki so v večvelikostnem primeru opremljene še z napovedjo relativne velikosti objekta, ne zgolj njegovega središča.

3.3.2.4 Povečanje robustnosti

Do sedaj smo predpostavili idealno invariantnost značilnic na spremembo velikosti, kar bi pomenilo, da povečevanje in zmanjševanje dane slike ohrani vse pojavitve značilnic na enakih relativnih položajih, velikosti pojavitev pa se spreminjajo premosorazmerno z velikostjo slike. V realnem svetu takšnih razmer ne moremo pričakovati, in sicer deloma zaradi neidealnosti obstoječih sestavov velikostno invariantnih značilnic, deloma pa zaradi izgubljanja informacije pri pomanjševanju slik.

V opisani shemi je možno povečati robustnost tako, da se v učni fazi, testni fazi ali pa kar v obeh fazah slike obravnavajo pri več kot eni velikosti. Pri večvelikostni obravnavi učnih slik ni treba izvesti nikakršnih prilagoditev. Pri večvelikostni obravnavi testnih slik pa je najenostavneje prevesti reprezentacije vseh velikostnih različic dane slike na velikost njene izvirne različice. Če se v reprezentaciji slike relativne velikosti v (glede na izvirno testno sliko) neka značilnica pojavi na položaju (x, y) in v relativni velikosti t , potem prevedba na izvirno velikost ta podatek pretvori v položaj $(x/v, y/v)$ in relativno velikost t/v . Po prevedbi je možno vse nastale reprezentacije iste slike združiti v eno samo množico podatkov o pojavitvah značilnic, ki se potem v detekcijskem postopku uporabi na običajen način.

3.4 Izhod detekcijskega postopka

Rezultat detekcijskega postopka je množica hipotez, ki napovedujejo položaje in velikosti domnevnih objektov v posameznih testnih slikah. Vrednotenje detekcijskih rezultatov med drugim zajema ugotavljanje pravilnosti hipotez, ki, kot bomo videli v naslednjem razdelku, temelji na primerjanju dejanskih in napovedanih omejevalnih okvirjev posameznih objektov. Zaradi tega mora detekcijski postopek za vsak hipotetični objekt izračunati njegov omejevalni okvir, ne zgolj njegovo središče in velikost.

Omejevalni okvir je določen s koordinatami središča in z velikostjo obeh stranic. Središče omejevalnega okvirja za dano hipotezo je seveda edino smiselno postaviti na središče objekta, ki ga hipoteza napoveduje. Podatek o velikosti okvirja pa je mogoče enostavno pridobiti na podlagi učnih podatkov in napovedi za relativno velikost hipotetičnega objekta. Pri enovelikostni detekciji je velikost omejevalnega okvirja najbolje postaviti kar na povprečno velikost učnih objektov. V naši im-

plementaciji so učni objekti, kot smo navedli, normalizirani bodisi glede na širino bodisi glede na višino omejevalnega okvirja. Tako so vsi učni objekti enako široki oziroma visoki, spreminja pa se njihova druga dimenzija.

Pri večvelikostni detekciji se velikost okvirja postavi na zmnožek napovedane relativne velikosti hipotetičnega objekta in povprečne velikosti objektov v izhodiščni učni velikosti. Pristop iz podrazdelka 3.3.1 predpostavlja, da se pravilna detekcijska hipoteza pojavi na plasti, pri kateri je velikost objekta približno enaka povprečni učni velikosti. Hipoteza, ki nastopa na plasti r piramide za testno sliko, napoveduje, da je velikost objekta v izvorni testni sliki enaka $2^{-r/R}$ -kratniku povprečne učne velikosti. Pristop iz podrazdelka 3.3.2 neposredno tvori hipoteze o relativnih velikostih posameznih objektov, zato hipoteza na plasti r glasovalnega prostora napoveduje, da je velikost objekta enaka $2^{r/R}$ -kratniku izhodiščne učne velikosti.

Izhod detekcijskega algoritma je tako množica detekcijskih hipotez. Vsaka hipoteza se nanaša na nek domnevni objekt, opisana pa je z naslednjimi podatki:

- Položaj središča omejevalnega okvirja domnevnega objekta, \mathbf{c} .
- Širina in višina omejevalnega okvirja.
- Moč (podpora) hipoteze, $s(H) \equiv s(\mathbf{c})$.

3.5 Vrednotenje rezultatov detekcije

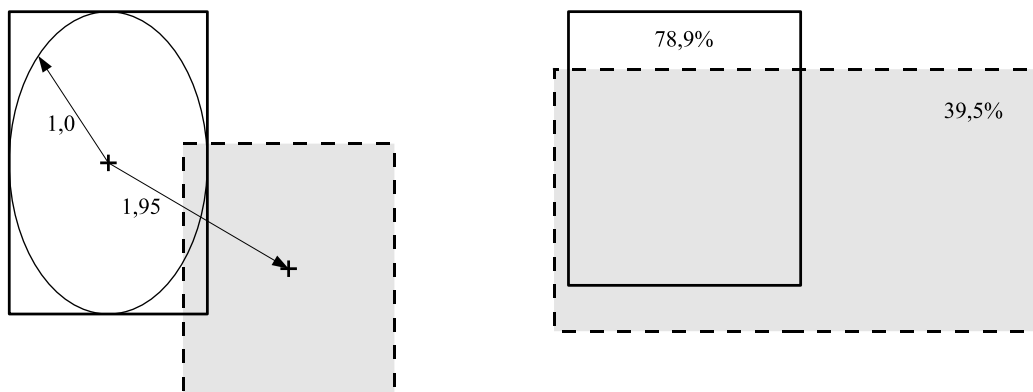
Kakovost detekcijskega postopka oziroma modela, ki ga postopek zgradi, je možno numerično ovrednotiti po preizkusu postopka oz. modela na dovolj veliki in raznoliki množici testnih slik. Detekcijski postopek za vsako testno sliko vrne množico detekcijskih hipotez. Ocena detekcijske točnosti je odvisna od pravilnosti posameznih hipotez in od detektorjeve stopnje zaupanja v posamezne hipoteze.

V podrazdelku 3.5.1 bomo predstavili odločitveno metodo za ugotavljanje pravilnosti posameznih hipotez. V podrazdelku 3.5.2 bomo opisali ocenjevanje točnosti detekcijskega modela po preizkusu na množici testnih slik.

3.5.1 Ugotavljanje pravilnosti detekcijskih hipotez

Kot je navedeno v razdelku 3.4, je vsaka hipoteza, ki jo za dano sliko proizvede detekcijski postopek, opremljena s podatki o omejevalnem okvirju domnevnega objekta, na katerega se hipoteza nanaša, in s količino $s(H)$, ki podaja detektorjevo stopnjo zaupanja v hipotezo.

Predpostavimo za začetek, da dana testna slika vsebuje natanko en objekt in da detekcijski postopek vrne natanko eno hipotezo. Naj (x_H, y_H) označuje središče omejevalnega okvirja, ki ga napoveduje detekcijska hipoteza, označi w_H in h_H pa naj po vrsti podajata širino in višino tega okvirja. (Do sedaj smo središče hipotetičnega okvirja označevali kot \mathbf{c} .) Naj x_O, y_O, w_O in h_O določajo dejanski



Slika 3.5: Ilustracija vrednotenja pogojev za preverjanje pravilnosti detekcijskih hipotez. Pravokotnika s polno obrobo predstavljata dejanska omejevalna okvirja dveh objektov, pravokotnika s črtkano obrobo pa napovedana omejevalna okvirja. Leva slika ilustrira računanje relativne razdalje d_r . Desna slika ilustrira koncept pokrivanja: napovedani okvir pokriva 78,9% dejanskega, dejanski okvir pa pokriva 39,5% napovedanega okvirja. (Slika je povzeta po Leibe (2004).)

omejevalni okvir (angl. ground truth) objekta na sliki. Potem hipoteza pravilno napoveduje položaj in velikost objekta, če veljajo vsi sledeči pogoji (Leibe, 2004):

1.

$$d_r = \sqrt{\left(\frac{2(x_H - x_O)}{w_O}\right)^2 + \left(\frac{2(y_H - y_O)}{h_O}\right)^2} \leq 0,5. \quad (3.20)$$

Količino d_r je možno razumeti na sledeč način. Naj bo E elipsa, včrtana dejanskemu omejevalnemu okvirju. Potem d_r predstavlja relativno razdaljo med dejanskim in napovedanim središčem objekta glede na „polmer“ elipse E (razdaljo od njenega središča do roba) na daljici, ki povezuje središči omejevalnih okvirjev. Če je (x_O, y_O) točka na elipsi E , velja $d_r = 1$. Dopustna vrednost $d_r = 0,5$ približno ustreza odstopanju napovedanega središča od dejanskega za četrtno dejanske velikosti objekta.

2. Napovedani omejevalni okvir pokriva najmanj 50% ploščine dejanskega omejevalnega okvirja.
3. Dejanski omejevalni okvir pokriva najmanj 50% ploščine napovedanega omejevalnega okvirja.

Slika 3.5 ilustrira vrednotenje navedenih pravilnostnih pogojev.

V splošnem lahko slika vsebuje poljubno število (vštevši nič) objektov, pa tudi detekcijski postopek lahko vrne poljubno število hipotez za sliko. V tem primeru se vsak napovedani omejevalni okvir primerja z vsakim dejanskim omejevalnim okvirjem po gornjih merilih. Poseben primer nastopi, če več kot ena hipoteza ustreza istemu objektu po teh merilih. V tem primeru se zgolj ena (katerakoli)

izmed ustrežajočih hipotez proglasi za pravilno, vse ostale ustrežajoče hipoteze pa se kljub izpolnjevanju navedenih pogojev obravnavajo kot napačne. Kot pravilno je smiselno označiti najmočnejšo ustrežajočo hipotezo, torej tisto z največjo vrednostjo $s(H)$. To merilo, ki kaznuje večkratno detektiranje istega objekta, so predlagali Agarwal in sod. (2004), upoštevajo ga pa tudi Leibe (2004) in drugi raziskovalci na področju detekcije objektov.

Predstavljene kriterije za ugotavljanje pravilnosti hipotez uporabljamo povsod razen pri preizkušanju detekcije na podatkovni bazi avtomobilov UIUC. V tem primeru smo se odločili za uporabo meril, ki so jih vzpostavili Agarwal in sod. (2004). Razlog za takšno odločitev je v želji po objektivnem primerjanju dobljenih rezultatov z rezultati, o katerih poročajo drugi raziskovalci, ki so svoje algoritme preizkušali na zbirki UIUC. Pri ostalih podatkovnih zbirkah ne izvajamo primerjav z drugimi avtorji, zato tam uporabljamo novejšje Leibejeve kriterije.

Po Agarwalu in sod. je hipoteza označena kot pravilna natanko tedaj, ko velja

$$\frac{(x_O - x_H)^2}{\alpha_h^2} + \frac{(y_O - y_H)^2}{\alpha_w^2} + \frac{(w_O - w_H)^2}{\alpha_w^2} \leq 1, \quad (3.21)$$

pri čemer je $\alpha_h = 0.25h_O$ in $\alpha_w = 0.25w_O$. Tudi po kriterijih Agarwala in sod. so večkratne detekcije istega objekta označene kot napačne.

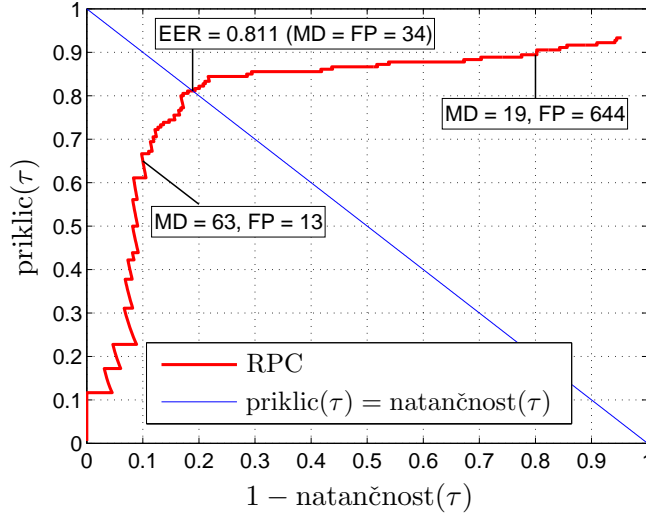
3.5.2 Ocenjevanje detekcijske točnosti

Vsaka hipoteza, ki jo vrne detekcijski postopek, je bodisi pravilna bodisi napačna glede na uporabljene kriterije. Lahko se tudi zgodi, da nekega prikazanega objekta ne napove nobena hipoteza. V angleški literaturi se pravilna detekcijska hipoteza pogosto označuje z izrazom *true positive*, napačna pa z izrazom *false positive*, kar bi lahko prevedli kot *lažni alarm*. Izraz *missing detection* (manjkajoča detekcija) predstavlja objekt na sliki, ki ga nobena hipoteza ne napove pravilno.

V nadaljevanju naj TP označuje število vseh pravilnih hipotez, ki jih detekcijski postopek proizvede za dano testno slikovno množico. Oznaka FP naj podaja število vseh napačnih hipotez. MD naj predstavlja število nedetektiranih objektov v celotni zbirki, OBJ pa skupno število objektov, ki jih vsebujejo testne slike. Velja torej $TP + MD = OBJ$. Meri *priklic* (angl. recall) in *natančnost* (angl. precision) sta definirani na sledeč način:

$$\begin{aligned} \text{priklic} &= \frac{TP}{OBJ} \\ \text{natančnost} &= \frac{TP}{TP + FP}. \end{aligned} \quad (3.22)$$

Najvišjo možno vrednost priklica ($\text{priklic} = 1$) je mogoče enostavno doseči tako, da detekcijski postopek vrne karseda gosto mrežo hipotez in vsem hipotezam dodeli isto vrednost $s(H)$. Na tak način bodo zanesljivo pravilno napovedani vsi objekti ($MD = 0$), vendar bo velika večina hipotez napačna, zato bo natančnost zelo majhna. Še enostavneje je doseči natančnost = 1: detekcijski



Slika 3.6: Primer krivulje RPC z oznakami MD in FP ter z označenim presečiščem s premico $\text{priklic}(\tau) = \text{natancnost}(\tau)$, ki opredeljuje mero EER. Za prikazano krivuljo RPC velja $\text{OBJ} = 180$.

postopek preprosto ne vrne nobene hipoteze. Seveda bo potem priklic enak 0. Priklic in natančnost sama po sebi očitno ne povesta ničesar o kakovosti detekcije. Veliko primernejša mera je krivulja RPC (angl. recall-precision curve), ki podaja odvisnost med priklicem in natančnostjo. Uporabo krivulje RPC pri vrednotenju detekcijske točnosti priporočajo Agarwal in sod. (2004). RPC je možno uporabiti le, če detekcijski postopek vsako hipotezo oceni s stopnjo zaupanja. Ta pogoj je v našem primeru seveda izpolnjen.

RPC je parametrično podana ravninska krivulja. Vlogo neodvisnega parametra igra prag τ , ki določa, katere detekcijske hipoteze, proizvedene za celotno testno množico, se upoštevajo pri računanju priklica in natančnosti. Pri vsaki vrednosti parametra τ se na podlagi vseh hipotez H , za katere velja $s(H) \geq \tau$, določita priklic in natančnost. Hipoteze, pri katerih velja $s(H) < \tau$, se pri računanju priklica in natančnosti za tekočo vrednost τ zanemarijo. Za vsak τ se izračunani par priklic-natančnost predstavi kot točka v koordinatni ravnini. Krivulja RPC se običajno prikaže kot odvisnost priklica od količine $(1 - \text{natancnost})$. Primer krivulje RPC je podan na sliki 3.6.

Mogoče je enostavno pokazati, da za izris popolne krivulje RPC zadošča, če prag τ preteče vrednosti $\{s(H)\}_{H \in \mathcal{H}} \cup \{\infty\}$, kjer \mathcal{H} označuje množico vseh detekcijskih hipotez za testne slike. Največji priklic in najmanjša natančnost bosta dosežena pri vrednosti $\tau = \min_{H \in \mathcal{H}} s(H)$, saj bodo v tem primeru upoštevane prav vse hipoteze. Pri $\tau = \infty$, ko pri izračunu priklica in natančnosti ne bo upoštevana nobena hipoteza, pa bo zanesljivo veljalo priklic = 0 in natančnost = 1. Pri povečevanju τ se v splošnem povečuje priklic in zmanjšuje natančnost.

Krivulja RPC je tem ugodnejša, čim bliže osema ($1 - \text{natančnost} = 0$) in (priklic = 1) poteka. Obstajata vsaj dve enostavni meri za določanje ugodnosti krivulje RPC:

Mera F_{\max} . To mero priporočajo Agarwal in sod. in je definirana kot

$$F_{\max} = \max_{\tau} \frac{2 \cdot \text{priklic}(\tau) \cdot \text{natančnost}(\tau)}{\text{priklic}(\tau) + \text{natančnost}(\tau)}. \quad (3.23)$$

Mera EER. Ta mera je definirana kot priklic (ali natančnost) v točki τ , kjer velja $\text{priklic}(\tau) = \text{natančnost}(\tau)$.

Meri predstavljata dva alternativna načina za iskanje kompromisa med priklicem in natančnostjo. Kot ilustrira slika 3.6, je mero EER mogoče določiti neposredno iz grafičnega prikaza krivulje, in sicer kot presečišče krivulje RPC s premico $\text{priklic}(\tau) = \text{natančnost}(\tau)$. Zaradi lažje berljivosti sta meri F_{\max} in EER pogosto podani v odstotkih.

V poglavju 5 bomo v nekaterih primerih prikazali celotno krivuljo RPC. Sicer bomo uspešnost detekcije vrednotili z obema merama, vendar bomo pogosteje prikazali le F_{\max} .

3.6 Povzetek poglavja in končne opombe

3.6.1 Povzetek poglavja

Poglavje smo namenili opisu detekcijske metode na osnovi implicitnega modela oblike in predstavitvi načina vrednotenja detekcijskih rezultatov.

Osnovni učni postopek predstavi obliko objektov dane kategorije z verjetnostno porazdelitvijo položaja središča objekta glede na položaje pojavitev posameznih značilnic v slikah. Porazdelitev se oceni na povsem neparametričen način.

Osnovni enovelikostni detekcijski postopek poveže podatke o položajih značilnic oz. lokalnih struktur v testni sliki z naučeno verjetnostno porazdelitvijo, tako da glasuje za hipoteze o možnih središčih objektov na sliki. Glasovi se zbirajo v dvorazsežnem hipotезnem prostoru. Končna množica hipotez se določi z iskanjem lokalnih maksimumov v zglajenem hipotезnem prostoru.

Pod določenimi pogoji, ki so pogosto izpolnjeni, je detekcijsko točnost mogoče znatno izboljšati z uporabo iterativnega postopka MDL, ki požrešno prireja lokalne strukture trenutno najmočnejšim hipotezam in posledično zmanjšuje moč konkurenčnim hipotezam.

Učni in testni postopek je možno razširiti za uporabo v večvelikostni detekcijski domeni, kjer pričakovani izhod detekcijskega algoritma obsega seznam hipotetičnih položajev in velikosti objektov na sliki. Predstavljeni sta bili dve različici za izvedbo večvelikostnega detekcijskega postopka. Po različici, ki deluje pri poljubnem sestavu značilnic, je učni algoritem enak osnovnemu učnemu algoritmu, testni algoritem pa zgradi piramido povečanih in pomanjšanih kopij dane testne

slike. Osnovni detekcijski postopek se nato požene na vsaki plasti piramide posebej. Na plasti, kjer se velikost nekega objekta na sliki približno izenači z znano učno velikostjo, obstaja možnost za nastop pravilne detekcijske hipoteze za ta objekt. Postopek MDL izbira in duši hipoteze vzdolž celotne piramide.

Različica izvedbe večvelikostnega detekcijskega postopka, ki predpostavlja invariantnost značilnic na spremembo velikosti, izkorišča dejstvo, da je na podlagi položajev in velikosti tovrstnih značilnic mogoče sklepati tudi na velikosti, ne zgolj na položaje objektov na sliki. Učni postopek zato oceni verjetnostno porazdelitev relativnega položaja središča objekta in njegove relativne velikosti glede na položaje in relativne velikosti nastopov posameznih značilnic na sliki. V testni fazi se ta informacija skupaj s položaji in relativnimi velikostmi manifestacij značilnic na dani testni sliki uporabi za glasovanje v trirazsežni hipotezni prostor. Končna množica hipotez se določi z glajenjem in iskanjem lokalnih maksimumov v tem prostoru. Ob koncu se moč sekundarnih hipotez zmanjša z uporabo algoritma MDL, ki poteka enako kot v enovelikostnem primeru.

Detekcijski modul poskrbi tudi za ugotavljanje pravilnosti hipotez, ki jih proizvede detekcijski algoritem, in za ocenjevanje točnosti detekcije objektov na slikovni množici.

3.6.2 Končne opombe

Detekcijski modul igra glavno vlogo v testni fazi, njegovih storitev pa se poslužujeta tudi obe izbirni metodi. V testni fazi se vedno uporablja popolna različica večvelikostnega detekcijskega postopka. V fazi izbire značilnic pa so razmere nekoliko drugačne. Pri računanju funkcije $S(Z, I)$ v okviru filtrske izbire značilnic ni smiselno izvajati postopka MDL, saj je funkcija odvisna le od najmočnejše detekcijske hipoteze. Pri izbiranju značilnic Fidler in sod. (razdelek 4.2), ki niso definirane na velikostno invarianten način, se uporablja le enovelikostna različica detekcijskega postopka.

Rezultati izvajanja detekcijskega postopka pa se v vseh primerih ovrednotijo na način, kot je bil predstavljen v razdelku 3.5.

Značilnice

4.1 Uvod

Z izbiro konkretnega sestava značilnic postane sistem v celoti opredeljen. Preizkusili smo dva sestava:

1. Sestav, predstavljen v člankih Fidler in sod. (2006) in Fidler in Leonardis (2007). Zaradi enostavnosti bomo ta sestav v nadaljevanju imenovali kar „značilnice Fidler in sod.“.
2. Sestav, predstavljen v člankih Ferrari in sod. (2006a) in Ferrari in sod. (2006b). Ta sestav bomo označevali kot „značilnice Ferrari in sod.“.

V obeh primerih gre za lokalne značilnice, ki jih je mogoče brez težav vključiti v izbirni in detekcijski modul. Oba sestava sta rezultat skrbnega in smotrnega načrtovanja, ki se odraža v dobro definirani zgradbi značilnic, v postopku učenja (pridobivanja) značilnic, ki sledi jasno začrtanim smernicam, ter v robustnem in učinkovitem postopku za reprezentacijo slik z naučenimi značilnicami. Oba sestava sta že bila uspešno preizkušena v kontekstih reprezentacije in detekcije objektov.

S stališča detekcijskega postopka se sestava značilnic bistveno razlikujeta. Značilnice Ferrari in sod. so namreč definirane na velikostno invarianten način, za značilnice Fidler in sod. pa to ne velja. Zato uporaba značilnic Ferrari in sod. omogoča detekcijo na podlagi velikostno invariantnega učnega modela (podrazdelek 3.3.2), značilnice Fidler in sod. pa so združljive le s postopkom iz podrazdelka 3.3.1.

Poudarjamo, da omenjena razlika v načinu definicije nikakor ne dopušča sklepanja, da bi bile značilnice Ferrari in sod. *a priori* „boljše“ od značilnic Fidler in sod. Neposredna primerjava obeh sestavov pravzaprav ni niti poštena niti smiselna, saj sta bila sestava razvita na podlagi precej različnih postulatov.

Vse postopke, povezane z značilnicami Fidler in sod., smo v celoti implementirali tudi sami. Pri tem smo se v nekaterih manj bistvenih vidikih nekoliko oddaljili od implementacije, opisane v obeh izvornih člankih, praviloma v smeri poenostavitve. Pri značilnicah Ferrari in sod. pa smo za pridobitev skupin robnih

odsekov iz dane slike, kar predstavlja osnovo za izgradnjo značilnic, uporabili prosto dostopno implementacijo¹. Pri končni fazi učenja značilnic in pri reprezentaciji slik pa smo uporabili lastno implementacijo.

Značilnice Fidler in sod. so predstavljene v razdelku 4.2, značilnice Ferrari in sod. pa v razdelku 4.3. Razdelek 4.4 povzema celotno poglavje.

4.2 Značilnice Fidler in sod.

4.2.1 Uvod

Fidler in sod. so razvili izviren sestav značilnic, ki temelji na hierarhični kompozicijski shemi. Celotna množica značilnic je organizirana v oštevilčene *ravni*. Vsaka značilnica ravni n , $n > 1$, je predstavljena kot geometrijska kompozicija ene ali več značilnic ravni $n - 1$. Izjema je raven 1, ki vsebuje osnovne, nedeljive značilnice, iz katerih so neposredno sestavljene vse značilnice ravni 2, posredno pa tudi značilnice višjih ravni. Značilnice v nižjih ravneh so preprostejše in manj distinktivne, tiste na višjih ravneh pa so kompleksnejše in bolj distinktivne. Fidler in sod. zaradi kompozicijske naravnosti sestava uporabljajo poimenovanje *del* (angl. part) za značilnico v hierarhični kompozicijski shemi.

Na prvi ravni hierarhije značilnic nastopajo enostavna usmerjena Gaborjeva jedra, ki so občutljiva na intenzitetne prehode (robove) v posameznih smereh. Vse nadaljnje ravni se zgradijo v postopku samodejnega učenja z obravnavo slik dane vizualne kategorije. Druga raven vsebuje pogosto nastopajoče lokalne kompozicije značilnic prve ravni, med katerimi je praviloma možno najti kote, polkroge in pare vzporednih črt. Značilnice na višjih ravneh v nekaterih primerih modelirajo posamezne razpoznavne dele objektov, na najvišjih pa celotne objekte oziroma kategorije. Konkretni primeri hierarhij značilnic so prikazani v obeh izvirnih člankih.

Kot je navedeno v Fidler in Leonardis (2007), so avtorji pri snovanju in izvedbi sestava značilnic upoštevali sledeče smernice:

1. Učinkovito *indeksiranje* (angl. indexing) značilnic v sliki in vizualnih kategorij na podlagi shranjenih (naučenih) značilnic. Značilnice na ravni n indeksirajo oz. služijo kot kazalo na posamezne značilnice na ravni $n + 1$. V smotrno organizirani hierarhični shemi posamezne značilnice indeksirajo le malo značilnic na višji ravni, kar omogoča učinkovito iskanje kompleksnejših značilnic in primerkov kategorij v dani sliki.
2. Učinkovito *primerjanje* (angl. matching) značilnic na sliki in naučenih značilnic. Tudi v tem pogledu se hierarhična kompozicijska predstavitev izkaže kot ustrezna odločitev.
3. Učenje, temelječe na statističnih lastnostih vizualnih podatkov. Fidler in sod. so predlagali učni postopek, ki gradi hierarhijo od spodaj navzgor,

¹<http://www.robots.ox.ac.uk/~ferrari/software.html>

usmerja pa ga pogostost pojavitev kandidatnih lokalnih kompozicij značilnic na obstoječih ravneh.

4. Odpornost na šum in variabilnost pri iskanju značilnic v sliki. V ta namen so značilnice predstavljene kot „ohlapne“ (angl. loose) geometrijske kompozicije svojih sestavnih delov, s čimer se skuša doseči kompromis med lokalno prostorsko povezanostjo značilnic in njihovo fleksibilnostjo.
5. Obstoj možnosti inkrementalnega učenja značilnic.

Podazdelek 4.2.2 natančneje opisuje zgradbo značilnic v hierarhični kompozicijski shemi. Podazdelek 4.2.3 je namenjen opisu postopka za reprezentacijo slik z dano množico značilnic. V podrazdelku 4.2.4 je predstavljen postopek učenja hierarhije značilnic. Tema podrazdelka 4.2.5 je uporaba značilnic Fidler in sod. v ostalih modulih sistema.

4.2.2 Predstavitev značilnic

4.2.2.1 Značilnica ravni 1

Prva raven vsebuje eno samo značilnico, označeno z Z_1^1 . Značilnica je definirana kot konvolucijsko jedro Gaborjevega filtra v orientaciji 0° . Kot bomo videli kasneje, pa lahko ta značilnica v sliki ali kot komponenta značilnic višjih ravni nastopa v katerikoli izmed osmih različnih orientacij. Množica možnih orientacij vsebuje vse cele večkratnike 45° med 0° in vključno 315° .

V implementaciji smo uporabili sledečo definicijo Gaborjevega filtra (Bovik in sod., 1990):

$$g(x, y) = \frac{1}{2\pi\lambda\sigma^2} \exp\left(-\frac{(x'/\lambda)^2 + y'^2}{2\sigma^2}\right) \sin(2\pi Fx'), \quad (4.1)$$

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

Parameter λ podaja razmerje polosi (angl. aspect ratio) eksponentne funkcije, σ pa njen razteg v ravnini xy . F podaja frekvenco sinusoidne funkcije, α pa njeno orientacijo. V naši implementaciji so vrednosti parametrov sledeče: $\lambda = 1$, $\sigma = 2,5$, $F = 2/35$ ($F\sigma = 1/7$) in $\alpha = 0^\circ$.

Po izračunu na dovolj veliki dvorazsežni domeni se funkcija $g(x, y)$ globalno normalizira tako, da vsota njenih vrednosti na domeni postane 0, vsota kvadratov vrednosti pa 1. Tako normalizirana funkcija (konvolucijsko jedro) se proglasi za edino značilnico ravni 1. Značilnica je prikazana na sliki 4.1.

4.2.2.2 Značilnice ravni $n > 1$

Vsaka značilnica je enolično določena z oznako ravni in indeksom znotraj ravni, ki skupaj tvorita *identitetno oznako*. Z_i^n označuje i -to značilnico na ravni n . Kot

smo že omenili, je vsaka značilnica ravni $n > 1$ sestavljena iz ene ali več značilnic ravni $n - 1$. *Kompozicija* značilnice Z_i^n je predstavljena kot

$$\text{kompozicija}(Z_i^n) = \langle (Z_{I(i,j)}^{n-1}, \alpha_{ij}, \mu_{ij}, \Sigma_{ij}) \rangle_j; \quad (4.2)$$

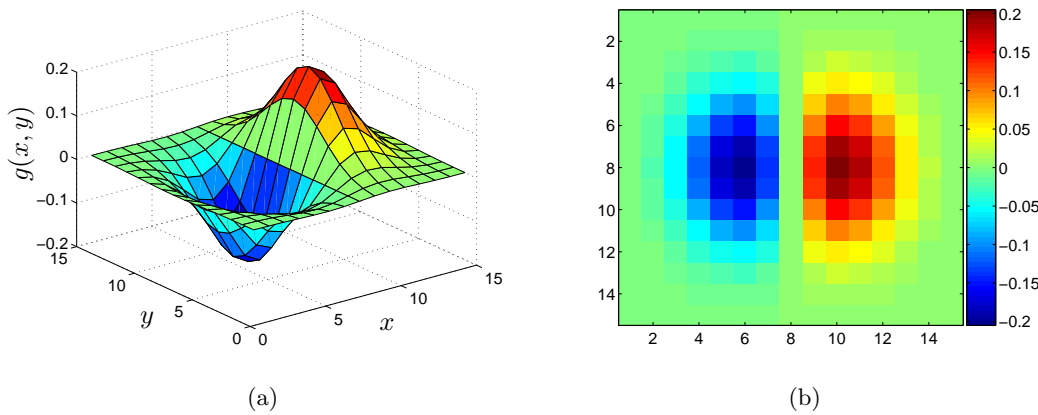
$I : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$ je indeksna funkcija.

Pri tem $Z_{I(i,j)}^{n-1}$ podaja identitetno oznako j -te komponente, α_{ij} njeno orientacijo (zavrtjenost), μ_{ij} in Σ_{ij} pa označujeta centroid in kovariančno matriko dvorazsežne Gaussove porazdelitve, ki modelira variabilnost položaja j -te komponente. Kompozicija je položajno in rotacijsko normalizirana glede na prvo komponento. Prva komponenta se po definiciji vedno nahaja v središču kompozicije in ima orientacijo enako 0° : $\alpha_{i1} = 0^\circ$, $\mu_{i1} = [0, 0]^T$, $\Sigma_{i1} = \mathbf{0}$. Položaji in orientacije ostalih komponent so izraženi relativno glede na prvo komponento. Vrednosti možnih orientacij so omejene na množico $\{(45k)^\circ\}_{k=0}^7$. Primer kompozicije značilnice je shematsko prikazan na sliki 4.2a. Tabela 4.1 prikazuje kompozicije treh primerov značilnic.

Kot bomo videli kasneje, kovariančna matrika Σ_{ij} posredno določa dopusten odmik komponente od relativnega položaja μ_{ij} . Na ta način se doseže kompromis med lokalno prostorsko povezanostjo značilnic in robustnostjo pri iskanju značilnic v sliki, kakor priporoča smernica 4 v podrazdelku 4.2.1.

4.2.3 Reprezentacija slik s podanimi značilnicami

Postopek reprezentacije dane slike z značilnicami ravni n poteka tako, da se slika najprej reprezentira z značilnicami ravni 1, nato z ravnjo 2 itn. Najprej bo predstavljen postopek reprezentacije izhodiščne slike z ravnjo 1, nato pa prehod z reprezentacije slike z ravnjo $n - 1$ na reprezentacijo z ravnjo n . Še pred tem pa bomo podrobneje opisali pojem manifestacije značilnice Fidler in sod.

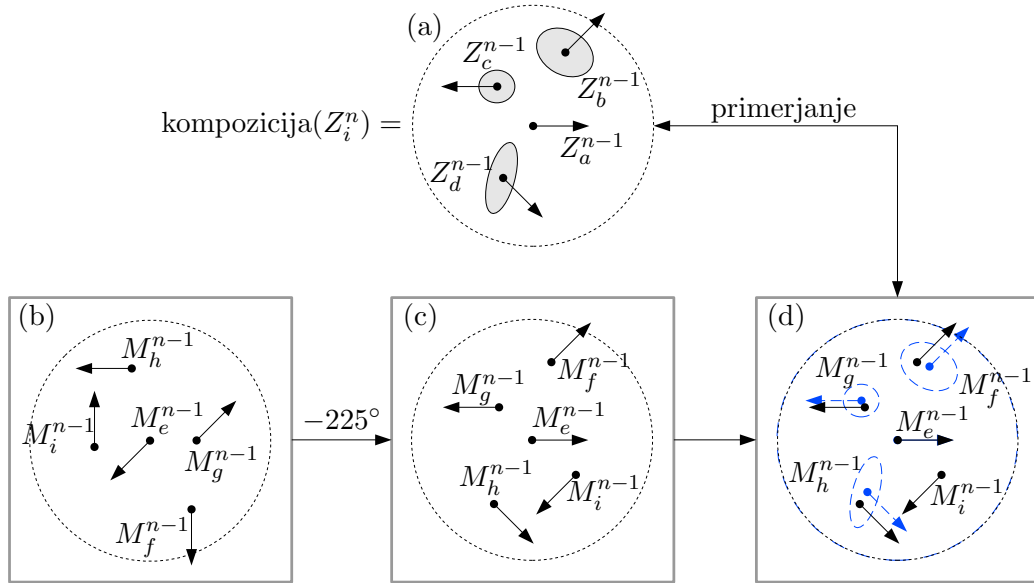


Slika 4.1: (a) Značilnica prve ravni kot funkcija dveh spremenljivk. (b) Ista značilnica, predočena kot matrika velikosti 15×15 , ki se uporablja kot konvolucijsko jedro pri reprezentaciji slik s prvo ravnjo.

4.2.3.1 Manifestacija značilnice




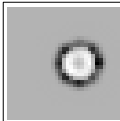

Pojem manifestacije (nastopa, pojavitve) značilnice smo uvedli že v uvodnem poglavju in ga nato vseskozi s pridom uporabljali. Kot smo navedli, *manifestacija* (angl. *realization* po Fidler in sod.) dane značilnice v sliki označuje posamezen odraz te značilnice v sliki. Reprezentacija slike s podanimi značilnicami je seznam manifestacij teh značilnic v obravnavani sliki. Z zapisom M_k^n bomo označili k -to manifestacijo v reprezentaciji slike z značilnicami ravni n .

V hierarhiji je vsaka značilnica predstavljena kot kompozicija, ki je normalizirana glede na prvo komponento. Koordinatno izhodišče shranjene značilnice je torej postavljeno v prvo komponento. V sliki pa lahko značilnica nastopa (se manifestira) na poljubnih položajih in v poljubnih orientacijah glede na slikovni koordinatni sistem. Denimo, da značilnica Z_i^n ($n > 1$) v sliki nastopa (tudi) kot manifestacija M_k^n , in sicer na položaju (x_k, y_k) in v orientaciji β_k . To dejstvo bomo zapisali $M_k^n = (Z_i^n, \beta_k, (x_k, y_k))$. Potem velja, da se prva komponenta značilnice nahaja na položaju (x_k, y_k) in v orientaciji β_k , ostale komponente pa se



Slika 4.2: (a) Kompozicija neke značilnice Z_i^n . Puščice označujejo orientacije komponent, elipse pa njihova dopustna odstopanja od centroidnih položajev. (b) Ali podoba prikazuje manifestacijo značilnice Z_i^n v orientaciji 225° na danem položaju v sliki? Denimo, da velja $\rho(e) = a$, $\rho(f) = b$, $\rho(g) = c$, $\rho(h) = d$. M_e^{n-1} torej indeksira Z_i^n . (c) Prvi korak algoritma za preverjanje nastopa značilnice Z_i^n : zavrtnitev okolice za kot -225° . (d) Algoritem preveri, če je za vse komponente značilnice mogoče najti ustrezajoče manifestacije. Z modro barvo so označene superpozicije komponent značilnice Z_i^n na okolico indeksirajoče manifestacije M_e^{n-1} . Izkaze se, da je (b) res manifestacija značilnice Z_i^n .

Tabela 4.1: Primeri značilnic za kategorijo avtomobili. Značilnica Z_1^1 , edina značilnica ravni 1, je predstavljena kot Gaborjevo jedro. Njena upodobitev je le pomanjšana sivinska različica slike 4.1b. Značilnica Z_{26}^2 je sestavljena iz treh premaknjenih in zavrtenih kopij Z_1^1 , značilnica Z_{95}^3 pa sestoji iz treh kopij značilnice Z_{26}^2 . V tabeli je razvidno, da prva komponenta vedno nastopa v izhodiščni orientaciji (0°) in v središču kompozicije.

Značilnica	Upodobitev	Kompozicija
Z_1^1		
Z_{26}^2		
Z_{95}^3		

glede na prvo komponento nahajajo v takšnih relativnih orientacijah in na takšnih relativnih položajih, kot to narekuje oz. dovoljuje kompozicija značilnice Z_i^n . Primer manifestacije značilnice s slike 4.2a je prikazan na sliki 4.2b.

Za poenostavitev izražanja bomo definirali funkcijo ρ , ki preslika indeks manifestacije znotraj reprezentacijskega seznama v indeks manifestirane značilnice. Za $M_k^n = (Z_i^n, \dots)$ torej velja $\rho(k) = i$.

4.2.3.2 Reprezentacija slike z ravnjo 1

Postopek se prične s konvoluiranjem slike z naborom osmih Gaborjevih jeder. Jedra so pridobljena na enak način kot značilnica Z_1^1 , le da se v enačbi (4.1) uporabijo drugačne vrednosti parametrov α , torej orientacij. Orientacija i -tega jedra znaša $45(i - 1)^\circ$. Prvo jedro je tako identično značilnici Z_1^1 .

Konvolucija za vsak slikovni element proizvede vektor odzivov posameznih jeder na vzorec v okolici tega elementa. Vsako jedro je občutljivo na intenzitetne prehode odgovarjajoče orientacije. Odziv jedra z orientacijo $\alpha + 180^\circ$ je v danem slikovnem elementu nasproten odzivu jedra orientacije α . Če slikovni element leži na vodoravnem prehodu temno-svetlo (levo temno, desno svetlo), bo največjo oz. najbolj pozitivno vrednost v njem doseglo jedro orientacije 0° , najmanjšo oz. najbolj negativno vrednost pa jedro orientacije 180° . Pri prehodu svetlo-temno bo stanje ravno obratno.

Vsakemu slikovnemu elementu se nato priredi indeks jedra z največjo vre-

dnostjo konvolucijskega odziva, ki je v nadaljevanju imenovano kar „najmočnejše jedro“. Tako je v izhodišču vsak slikovni element manifestacija značilnice Z_1^1 v orientaciji najmočnejšega jedra.

V naslednjem koraku se ohranijo le tisti slikovni elementi, pri katerih odziv najmočnejšega jedra presega $1/10$ vrednosti globalno največjega odziva v sliki. Na ta način se iz nadaljnje obravnave izločijo homogena območja, kar v splošnem ne predstavlja posebne škode, saj je večina relevantne informacije zajete v intenzitetnih prehodih.

Jedra ustreznih orientacij praviloma zanesljivo zaznajo usmerjene intenzitetne prehode. Najmočnejše se odzovejo v samih prehodih, nekoliko šibkeje pa tudi v slikovnih elementih v neposredni bližini prehodov. Postopek *lokalnega dušenja* (angl. local suppression) ohrani le manifestacije jeder na samih prehodih, ne pa tudi ob njih, s čimer se prihrani precej pomnilniškega prostora. Postopek sprejme matriko vrednosti odzivov najmočnejših jeder v posameznih slikovnih elementih in za vsak slikovni element ugotovi, koliko vrednosti odzivov v njegovi okolici velikosti 3×3 je večjih ali enakih kot vrednost odziva v obravnavanem elementu. Če je število takih vrednosti enako vsaj 3, se obravnavani element odstrani iz reprezentacije.

4.2.3.3 Prehod z reprezentacije slike z ravňjo $n - 1$ na reprezentacijo z ravňjo n

Predpostavimo, da je slika že reprezentirana z značilnicami ravni $n - 1$, torej z množico manifestacij $\{M_k^{n-1}\}_k$, $M_k^{n-1} = (Z_{\rho(k)}^{n-1}, \beta_k, (x_k, y_k))$. V nadaljevanju bomo opisali, kako je na podlagi teh podatkov možno zgraditi reprezentacijo slike z značilnicami ravni n .

Postopek za prehod na reprezentacijo ravni n obravnava vsako manifestacijo M_k^{n-1} posebej. Najprej poišče vse značilnice Z_i^n , pri katerih se identiteta prve komponente ujema z identiteto manifestirane značilnice, torej $Z_{I(i,1)}^{n-1} = Z_{\rho(k)}^{n-1}$. Pravimo, da značilnica $Z_{\rho(k)}^{n-1}$ oziroma njena manifestacija *indeksira* vse takšne značilnice Z_i^n . Postopek nato za vsako indeksirano značilnico Z_i^n preveri, ali se v celoti, z vsemi komponentami, manifestira na položaju (x_k, y_k) in v orientaciji β_k . Algoritem za preverjanje nastopa značilnice Z_i^n na mestu dane manifestacije je ključna sestavina reprezentacijskega postopka, zato ga izpostavljamo kot algoritem 4.1 in ilustriramo na sliki 4.2.

Algoritem najprej zavrti okolico M_k^{n-1} okrog točke (x_k, y_k) za kot $-\beta_k$, s čimer orientacija M_k^{n-1} postane 0° . Na ta način se M_k^{n-1} „poravna“ s prvo komponento indeksirane značilnice Z_i^n , kar omogoči neposredno primerjavo ostalih komponent Z_i^n z manifestacijami v okolici M_k^{n-1} . Algoritem vrne odgovor „DA“ tedaj, ko je za vse komponente Z_i^n možno najti ujemajoče manifestacije v zavrti okolici M_k^{n-1} . Komponenta značilnice se ujema z manifestacijo, če se ujemata identiteti in orientaciji, relativni položaj manifestacije glede na položaj M_k^{n-1} pa ne odstopa pretirano od centroidnega položaja komponente. Dopustno odstopanje določa kovariančna matrika komponente.

Algoritem 4.1 Preverjanje nastopa značilnice Z_i^n na položaju in v orientaciji dane manifestacije M_k^{n-1} v reprezentaciji slike z značilnicami ravni $n - 1$.

Vhod:

- (a) Seznam manifestacij $\{M_l^{n-1}\}_l$ (reprezentacija slike z ravno $n - 1$).
- (b) Manifestacija $M_k^{n-1} = (Z_{\rho(k)}^{n-1}, \beta_k, (x_k, y_k))$.
- (c) Značilnica $Z_i^n = \langle (Z_{I(i,j)}^{n-1}, \alpha_{ij}, \mu_{ij}, \Sigma_{ij}) \rangle_j$, tako da velja $Z_{I(i,1)}^{n-1} = Z_{\rho(k)}^{n-1}$.

Izhod:

„DA“, če se Z_i^n manifestira na mestu M_k^{n-1} ; „NE“ sicer.

Postopek:

- 1: Zavrti (dovolj veliko) prostorsko okolico M_k^{n-1} za kot $-\beta_k$ okrog (x_k, y_k) . Tako se manifestacija M_k^{n-1} pretvori v $(Z_{\rho(k)}^{n-1}, 0^\circ, (x_k, y_k))$, ustrezno pa se spremeni tudi položaji in orientacije drugih manifestacij v okolici.
 - 2: **za** $j := 2$ **do** $\# \text{komponent}(Z_i^n)$ **ponavlja**
 - 3: Preveri, ali zavrtena okolica vsebuje manifestacijo $M_r^{n-1} = (Z_{\rho(r)}^{n-1}, \beta_r, (x_r, y_r))$ (elementi vektorja odražajo stanje po vrtenju okolice), za katero so izpolnjeni vsi sledeči pogoji:
 - (a) $Z_{I(i,j)}^{n-1} = Z_{\rho(r)}^{n-1}$;
 - (b) $\alpha_{ij} = \beta_r$;
 - (c) $\exp(-\frac{1}{2}(\mathbf{x} - \mu_{ij})^T \Sigma_{ij}^{-1}(\mathbf{x} - \mu_{ij})) \geq 0,2$, kjer $\mathbf{x} = [x_r - x_k, y_r - y_k]^T$.
 - 4: če takšna manifestacija M_r^{n-1} **ne** obstaja, **potem**
 - 5: **vrni** „NE“ // *izvajanje se zaključ*
 - 6: **konec** {če}
 - 7: **konec** {za}
 - 8: **vrni** „DA“
-

Za vsako značilnico Z_i^n , ki se manifestira na mestu M_k^{n-1} , se v reprezentacijski seznam za raven n doda element $(Z_i^n, \beta_k, (x_k, y_k))$. Po določitvi ustrežajočih značilnic ravni n za vse manifestacije ravni $n - 1$ je slika povsem reprezentirana z ravno n .

Kot je navedeno v podrazdelku 4.2.4, se po koncu učenja ravni $n \geq 3$ izvedeta postopka gručenja (angl. clustering) in zmanjšanja prostorske ločljivosti. Posledica gručenja je preimenovanje izvornih identitetnih oznak značilnic v oznake gruč. Podatki, potrebni za izvedbo te transformacije, se shranijo v učni fazi skupaj z izvornimi in transformiranimi značilnicami (t.j. gručami). Pri prehodu z ravni $n - 1$ na raven n ($n \geq 3$) reprezentacijski postopek najprej poišče manifestacije izvornih značilnic, kot je to opisano zgoraj, nato pa v skladu s transformacijo preimenuje identitetne oznake značilnic v odkritih manifestacijah.

Zmanjšanje prostorske ločljivosti se odraža v spremembi položajnih koordinat manifestacij za konstanten faktor. Tudi to se izvrši le v primeru $n \geq 3$.

4.2.4 Učenje hierarhije značilnic

Postopek učenja, ki ga predstavljamo v tem podrazdelku, gradi hierarhijo značilnic od spodaj navzgor. Raven 1 je določena vnaprej, raven $n > 1$ pa se zgradi s statističnim učenjem na podlagi reprezentacij učnih slik z ravno $n - 1$.

4.2.4.1 Vhod

Postopek učenja ravni n sprejme množico učnih slik, reprezentiranih z značilnicami ravni $n - 1$. Izkaže pa se, da je pred učenjem koristno dodatno zmanjšati število manifestacij v reprezentacijski množici z uporabo postopka *inhibicije*.

Inhibicijo reprezentacije z značilnicami prve ravni najlažje razumemo, če si predstavljamo intenzitetni prehod, denimo, tipa „zgoraj temno, spodaj svetlo“, na katerega se najmočnejše odzove jedro orientacije 90° . Po lokalnem dušenju so manifestacije tega jedra nanizane v sklenjeni vodoravni črti na prehodu. Če se vsaka manifestacija na črti nadomesti s sliko uporabljenega 90-stopinjskega Gaborjevega jedra, pride do precejšnjega prekrivanja, kar nakazuje določeno redundanco predstavitve. Črto bi bilo možno enako kakovostno pokriti z bistveno manj jedri oziroma manifestacijami. Cilj postopka inhibicije, ki so ga Fidler in sod. (2006) formulirali v ogrodju načela MDL, je izločiti čimveč manifestacij, pri tem pa upoštevati zahtevo o „dovolj dobri“ pokritosti intenzitetnih prehodov z Gaborjevimi jedri. Koncept pokritosti in inhibicijski postopek je možno opredeliti tudi na višjih ravneh (manifestacija značilnice pokriva manifestacije komponent te značilnice).

Inhibicijo je možno izvesti s preprostim požrešnim postopkom, ki precej spominja na postopek MDL pri detekciji. Inhibicija se uporablja le pri učenju hierarhije značilnic, in sicer iz razlogov prostorsko-časovne narave, pri reprezentaciji slik z že zgrajeno hierarhijo značilnic pa se inhibicija ne uporablja, saj zmanjšuje možnosti za odkrivanje manifestacij kompleksnejših značilnic.

4.2.4.2 Učenje kompozicij

Pri učenju množice značilnic dane ravni $n > 1$ se število komponent posameznih značilnic postopoma povečuje: najprej se zgradijo značilnice, ki sestojijo iz središčne (prve) komponente in še ene značilnice ravni $n - 1$, nato značilnice z dvema komponentama poleg središčne itn. Kompozicije, ki poleg središčne značilnice vsebujejo še natanko s komponent, bomo imenovali s -kompozicije (angl. *s-subcompositions* po Fidler in sod.). Najprej bomo opisali učenje 1-kompozicij, nato pa učenje s -kompozicij pri poljubnem $s > 1$.

Izhodišče za gradnjo 1-kompozicij ravni n je množica inhibiranih reprezentacij učnih slik z značilnicami ravni $n - 1$. Na začetku učenja se za vsako trojico $(Z_i^{n-1}, Z_j^{n-1}, \alpha)$ — takih trojic je $V(n-1) \times V(n-1) \times 8$, če $V(\cdot)$ označuje število značilnic dane ravni — pripravi *prostorski zemljevid* (angl. spatial map), ki bo tekom obravnave reprezentiranih slik zbiral informacijo o tem, kje in kako pogosto se v prostorski okolici manifestacij značilnic Z_i^{n-1} pojavljajo manifestacije

značilnic Z_j^{n-1} v relativni orientaciji α . V naši implementaciji ima okolica obliko kroga premera 21 slikovnih elementov, sicer pa je postopek za določitev najprimernejše velikosti okolice opisan v Fidler in sod. (2006). Prostorski zemljevid je tako inicializiran na ničelno matriko velikosti 21×21 .

Učni postopek nato sistematično pregleda vhodne reprezentacije in obravnava vsako manifestacijo $M_k^{n-1} = (Z_{\rho(k)}^{n-1}, \beta_k, (x_k, y_k))$ posebej. Okolica obravnavane manifestacije se najprej zavrti za kot $-\beta_k$ okrog (x_k, y_k) , s čimer se ustrezno spremenijo parametri vseh manifestacij v okolici. Nato postopek pregleda vsebino celotne zavrtene okolice. Za vsako manifestacijo $(Z_{\rho(l)}^{n-1}, \beta_l, (x_l, y_l))$, ki se nahaja v zavrti okolici, postopek prišteje 1 v celico $(x_l - x_k, y_l - y_k)$ zemljevida za trojico $(Z_{\rho(k)}^{n-1}, Z_{\rho(l)}^{n-1}, \beta_l)$.

Po obravnavi vseh učnih reprezentacij se vsebina vsakega zemljevida aproksimira z linearno kombinacijo Gaussovih gruč. Vsaka gruča se opiše s centroidom (μ), kovariančno matriko (Σ) in skupnim številom glasov, ki je enako vsoti vrednosti celic, ki pripadajo gruči. Ker so posamezne gruče pogosto precej jasno ločene od ozadja in drugih gruč, se njihovi parametri lahko določijo z enostavnimi heurističnimi postopki.

V naslednjem koraku se za vsako Gaussovo gručo vsakega zemljevida tvori po ena 1-kompozicijska značilnica. Zemljevid za trojico $(Z_i^{n-1}, Z_j^{n-1}, \alpha)$, ki vsebuje gruče $\{(\mu_k, \Sigma_k)\}_k$, prispeva množico značilnic

$$\{ \langle (Z_i^{n-1}, 0^\circ, [0, 0]^T, \mathbf{0}), (Z_j^{n-1}, \alpha, \mu_k, \Sigma_k) \rangle \}_k$$

v seznam naučenih 1-kompozicij. Da gradnja 2-kompozicij ne bi presegla omejitev računalnikovih virov, se ohrani do 200 najmočnejše zastopanih 1-kompozicij. Kot mera zastopanosti služi število glasov, ki so jih prejele gruče, iz katerih izvirajo 1-kompozicije. Fidler in Leonardis (2007) sta predlagala kompleksnejši in teoretsko boljše podprt izbirni mehanizem.

Izhodišče za gradnjo s -kompozicij pri $s > 1$ sestoji iz množice naučenih $(s-1)$ -kompozicij ravni n in množice vhodnih reprezentacij ravni $n-1$. Za vsako trojico $(Z_i^{n,s-1}, Z_j^{n-1}, \alpha)$, kjer $Z_i^{n,s-1}$ označuje i -to naučeno $(s-1)$ -kompozicijo, se pripravi enak prostorski zemljevid kot pri gradnji 1-kompozicij. Algoritem nato išče naučene $(s-1)$ -kompozicije v vhodnih reprezentacijah z uporabo običajnega reprezentacijskega postopka. Kadar na nekem položaju nastopa značilnica $Z_i^{n,s-1}$, v okolici tega položaja pa poleg komponent $Z_i^{n,s-1}$ še (vsaj) značilnica Z_j^{n-1} v relativni orientaciji β_j glede na manifestacijo $Z_i^{n,s-1}$, učni postopek prišteje 1 v ustrezno celico zemljevida za trojico $(Z_i^{n,s-1}, Z_j^{n-1}, \beta_j)$. Po obravnavi vseh učnih reprezentacij se izvede enak zaključni postopek kot v primeru 1-kompozicij: na podlagi vsebin zemljevidov se zgradijo značilnice z s komponentami poleg središčne, nato pa se ohrani 200 najmočnejše zastopanih značilnic.

Postopek učenja ravni n se v naši implementaciji ustavi po zaključku gradnje 3-kompozicij.

4.2.4.3 Gručenje značilnic

Učni postopek lahko po različnih poteh zgradi podobne ali celo enake značilnice. Ker manjše število značilnic poveča verjetnost nastopa statistično dobro podprtih kompozicij pri učenju naslednje ravni, je smotrno združiti dovolj podobne značilnice v gruče. Vendar pa je z direktnim primerjanjem kompozicijskih zapisov težko prepoznati, kako podobni sta si dani značilnici. V implementaciji zato izvajamo gručenje na podlagi podobnosti upodobitev značilnic dane ravni. Ker se ta pristop bistveno razlikuje od pristopa Fidler in Leonardis (2007), mu namenimo nekoliko več pozornosti.

Upodobitev značilnice Z_1^1 je definirana kot intenzitetna slika Gaborjevega jedra v orientaciji 0° . Upodobitev značilnice Z_i^n s kompozicijo $\langle (Z_{I(i,j)}^{n-1}, \alpha_{ij}, \mu_{ij}, \Sigma_{ij}) \rangle_j$ pa je definirana kot superpozicija upodobitev komponent, pri čemer je upodobitev značilnice $Z_{I(i,j)}^{n-1}$ zavrtena za kot α_{ij} in postavljena na relativen položaj μ_{ij} . Primeri upodobitev so prikazani v tabeli 4.1.

Za mero podobnosti med danima upodobitvama značilnic, ki usmerja postopek gručenja, se uporablja normalizirana sivinska korelacija (angl. *normalized gray-scale correlation*):

$$\text{NGC}(A, B) = \frac{\sum_i \sum_j (A_{ij} - \bar{A})(B_{ij} - \bar{B})}{\sqrt{\sum_i \sum_j (A_{ij} - \bar{A})^2 \sum_i \sum_j (B_{ij} - \bar{B})^2}}, \quad (4.3)$$

kjer A in B označujeta intenzitetni sliki enake velikosti, A_{ij} in B_{ij} posamezne slikovne elemente, \bar{A} in \bar{B} pa povprečno intenziteto slik A in B . NGC zavzema vrednosti med -1 in 1 . Ker so vse upodobitve značilnic iste ravni enako velike in ker se upodobitev središčne komponente značilnice vedno nahaja v središču upodobitve celotne značilnice, je uporaba mere NGC smiselna.

Postopek gručenja značilnic dane ravni deluje po načelu „od spodaj navzgor“ in se zato označuje tudi z imenom *aglomerativno gručenje*. V začetku vsaka značilnica predstavlja svojo gručo. Nato se v vsaki iteraciji združita dve najbolj podobni gruči. Postopek se ponavlja, dokler je podobnost med najbolj podobnima gručama nad določenim pragom. Podobnost med gručama C_1 in C_2 je možno definirati na več načinov (Lance in Williams, 1967). V implementaciji je opredeljena kot

$$\text{podobnost}(C_1, C_2) = \min_{A \in C_1, B \in C_2} \text{NGC}(A, B). \quad (4.4)$$

Podobnostni prag je v implementaciji postavljen na 0,85. V okolju Matlab je takšno gručenje najlažje izvesti s kombinacijo funkcij `linkage` in `cluster`.

Po zaključku gručenja se izvrši preimenovanje značilnic dane ravni, saj vsaka gruča postane nova značilnica. Identitetna oznaka vsake izvirne značilnice se spremeni v oznako gruče, ki ji značilnica pripada. Na ravni $n + 1$ ni več mogoče razlikovati med značilnicami iste gruče ravni n .

Upodobitev gruče oz. novonastale značilnice je definirana kot povprečna slika upodobitev značilnic iz gruče.

4.2.4.4 Zmanjšanje prostorske ločljivosti

Ker značilnice z vzpenjanjem po hierarhiji postajajo čedalje kompleksnejše, je smiselno, da pokrivajo čedalje večja slikovna območja. Največja možna velikost območja, ki ga pokrivajo značilnice dane ravni, je določena s premerom krožne okolice, ki se uporablja v fazi polnjenja prostorskih zemljevidov. Potemtakem bi se velikost okolice morala povečati pred začetkom učenja tretje in vsake nadaljnje ravni.

Fidler in Leonardis (2007) predlagata drugačno, časovno učinkovitejšo rešitev. Velikost okolice naj v absolutnem smislu ostane konstantna, pač pa naj se zmanjšuje prostorska ločljivost reprezentacij. V naši implementaciji se pred pričetkom učenja vsake ravni od $n = 3$ naprej prostorska ločljivost spremeni za faktor $11/21$, kar tehnično pomeni množenje vseh položajnih vektorjev (x_k, y_k) v reprezentacijskem seznamu za ta faktor. Velikost okolice oz. prostorskih zemljevidov vseskozi ostaja 21×21 enot, le ploščina enote napram ploščini slikovnega elementa izvirne slike se poveča za faktor $(21/11)^2$ pred pričetkom gradnje tretje in vsake nadaljnje ravni.

4.2.5 Uporaba značilnic Fidler in sod. v drugih moduli sistema

Kot smo navedli, so značilnice Fidler in sod. v hierarhiji predstavljene na rotacijsko normaliziran način. Značilnica predstavlja ekvivalenčni razred vseh kompozicij, ki se razlikujejo zgolj po zavrtenosti okrog prve (središčne) komponente. Za uporabo v izbirnem in detekcijskem modulu sistema pa je bolj pripravno, če značilnice niso rotacijsko normalizirane, saj detekcijski postopek ni rotacijsko invarianten. Zato se v teh moduli različno zavrtene kompozicije obravnavajo kot različne značilnice. Značilnic, kot jih pojmuje v izbirnem modulu, je potemtakem osemkrat več kot značilnic, kot jih pojmuje v fazi učenja hierarhije — vsaka zgrajena značilnica se razmnoži v osem različic, po eno za vsako orientacijo.

Reprezentacija slike z značilnicami Fidler in sod. je vselej enolična. Verjetnost $P(G|L, \mathbf{p})$, ki nastopa v enačbi (3.2) in tudi kasneje v poglavju 3, tako znaša 1, če se značilnica G manifestira v lokalni strukturi L na položaju \mathbf{p} , sicer pa velja $P(G|L, \mathbf{p}) = 0$. Naj opozorimo, da na istem položaju v sliki lahko nastopa več različnih značilnic, vendar te tehnično pripadajo različnim lokalnim strukturam L . Zato pri vsakem paru (L, \mathbf{p}) obstaja kvečjemu ena značilnica G , tako da velja $P(G|L, \mathbf{p}) = 1$; za vse druge značilnice velja $P(G|L, \mathbf{p}) = 0$.

Zaradi velikostne neinvariantnosti značilnic Fidler in sod. se večvelikostna detekcija lahko izvaja le po postopku iz podrazdelka 3.3.1.

4.3 Značilnice Ferrari in sod.

4.3.1 Uvod

Metoda za gradnjo in uporabo značilnic Ferrari in sod. izkorišča informacijo, zajeto v intenzitetnih prehodih (robvih) v danih slikah. Na takšni podlagi temelji

tudi metoda Fidler in sod., vendar se pristopa kljub podobnim izhodiščem precej razlikujeta. Značilnice Ferrari in sod. v nasprotju z značilnicami Fidler in sod. niso organizirane v hierarhično shemo in niso predstavljene kot kompozicija drugih značilnic.

Pri metodi Fidler in sod. je gonilna sila za učenje značilnic statistika manifestacij osnovnih in izpeljanih značilnic v danih vizualnih podatkih. Metoda Ferrari in sod. pa gradi značilnice na osnovi *povezanosti* robnih elementov v učnih slikah. V prvi fazi se robni elementi (angl. *edgels*) združujejo v verige. Verige se nato razdelijo na približno ravne odseke, ki se v naslednji fazi združujejo v skupine. V obeh združevalnih fazah je ključni kriterij za združevanje povezanost.

Vsaka značilnica Ferrari in sod. zajema gručo podobnih *skupin robnih odsekov*. Vsako skupino sestavlja k približno ravnih odsekov, ki so med seboj povezani (angl. *k adjacent segments, kAS*). Pri $k = 1$ nabor značilnic tvorijo kar posamezni odseki različnih orientacij; pri $k = 2$ nabor med drugim vsebuje značilnice v obliki črk T in L; nabor za $k = 3$ vključuje značilnice v obliki črk C, Z in F; in tako naprej. Z večjim k se distinktivnost posameznih značilnic v splošnem povečuje, pogostost in ponovljivost njihovega nastopanja pa se zmanjšujeta. Brez izgube splošnosti lahko predpostavimo, da je parameter k določen vnaprej in vseskozi enak. V nasprotnem primeru se učenje značilnic in reprezentiranje slik izvajata za vsak k posebej.

Opisniki značilnic so invariantni v položajnem in v velikostnem smislu. V praksi se pri velikostni invariantnosti sicer pokažejo določene omejitve, vendar kljub tem pomanjkljivostim v fazi detekcije privzemamo invariantnost značilnic na spremembo velikosti, kar nam omogoča uporabo postopka iz podrazdelka 3.3.2.

Podrazdelek 4.3.2 predstavlja postopke pridobivanja, opisovanja in primerjanja skupin robnih odsekov. V podrazdelku 4.3.3 je opisan postopek gradnje značilnic na podlagi skupin odsekov, pridobljenih iz učnih slik. Tema podrazdelka 4.3.4 je reprezentacija slik s podanimi značilnicami. Podrazdelek 4.3.5 predstavi značilnice Ferrari in sod. v povezavi s celotnim sistemom.

4.3.2 Skupine k robnih odsekov

Za pridobivanje in opisovanje (ne pa tudi za primerjanje) skupin robnih odsekov smo uporabili kar prosto dostopno implementacijo Ferrarija in sod. Zato sta sledeča odseka v celoti povzeta po Ferrari in sod. (2006a).

4.3.2.1 Pridobivanje skupin

Postopek za pridobivanje skupin robnih odsekov se prične z detekcijo robov, čemur služi t.i. detektor Berkeley (Martin in sod., 2004). V naslednji fazi postopek poveže sosednje robne elemente združljivih orientacij v verige. Zaradi nepopolnosti detektorja robov se intenzitetni prehod, ki je v izvorni sliki videti sklenjen, lahko pretvori v več nepovezanih verig. To se pogosto zgodi ob presečiščih robov. Postopek premosti dovolj majhne presledke med verigami, ki najverjetneje pripadajo istemu sklenjenemu robu, s čimer pridobivanje odsekov robnih elementov

postane robustnejše. To se pokaže predvsem pri reprezentaciji slik z že naučenimi značilnicami.

V sledečem koraku se vsaka dobljena veriga, ki ne zadošča določenemu kriteriju ravnosti, razbije na povezano skupino približno ravnih odsekov. Postopek nato uredi nastale odseke v neusmerjen graf, imenovan *mreža robnih odsekov* (angl. contour segment network, CSN). Vozlišča v mreži ustrezajo posameznim odsekom, povezave med vozlišči pa so določene z množico hevrističnih pravil. V skladu s pravili se odseka, ki se na sliki dotikata, preslikata v par povezanih vozlišč v mreži. Vendar pa je v skladu s hevrističnimi pravili par vozlišč v mreži lahko povezan, tudi če se ustrezna odseka na sliki „skoraj“ dotikata. Pojma „dotikanje“ in „skorajšnje dotikanje“ vključujeta tudi dotikanje v smislu L- in T-presečišč ter podobnih struktur, ne zgolj dotikanje v smislu nadaljevanja verige.

Skupine k robnih odsekov so definirane preprosto kot poti dolžine k v mreži povezanih odsekov. Takšne poti se iz mreže pridobijo tako, da se v vsakem vozlišču požene postopek iskanja v globino, nato pa se za vsako množico odsekov ohrani le ena pot. Lahko se namreč zgodi, da iskalni algoritem proizvede več različnih poti, ki pa so sestavljene iz istih odsekov.

4.3.2.2 Opisovanje skupin

Opisnik dane skupine k robnih odsekov je numerični vektor, ki vsebuje podatke o relativnih položajih, relativnih dolžinah in absolutnih orientacijah posameznih odsekov v skupini. Z definicijo mere razdalje, kar je tema naslednjega podrazdelka, se problem primerjanja skupin prevede na problem primerjanja numeričnih opisnikov.

Pri opisovanju se odseki aproksimirajo kar z daljicami. Ta korak je smiseln in utemeljen, saj so odseki skupin že po načinu pridobivanja približno ravni. Za vsak odsek se nato izračuna središče aproksimirajoče daljice. Množica središč je označena kot $\{\mathbf{M}_i\}_{i=1}^k$, $\mathbf{M}_i = (x_i, y_i)$ pa podaja posamezno središče daljice.

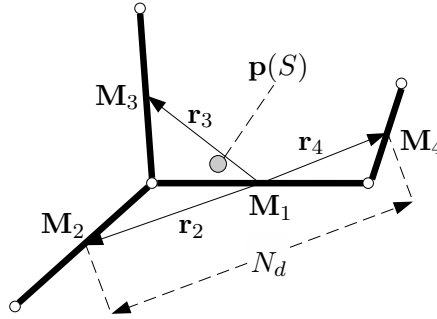
V opisnem vektorju skupine so podatki o posameznih odsekih navedeni po nekem vrstnem redu. Zaželeno je, če je vrstni red stabilen in robusten, tako da so odseki v vizualno podobnih skupinah vedno razvrščeni na enak način. Ferrari in sod. dodelijo prvo mesto v vrstnem redu tistemu odseku, čigar središčna točka \mathbf{M}_i leži najbližje težišču (centroidu) središčnih točk vseh odsekov. Ostali odseki se razvrstijo od leve proti desni glede na položaje njihovih središčnih točk. Če je razlika med koordinatami x središčnih točk dane podmnožice odsekov premajhna za zanesljivo razvrstitev, potem se taka podmnožica uredi od zgoraj navzdol, torej glede na koordinate y središčnih točk. Primer urejene skupine robnih odsekov je prikazan na sliki 4.3.

Naj $S = (s_1, \dots, s_k)$ označuje skupino k robnih odsekov. Zapis $\mathbf{r}_i = (r_i^x, r_i^y)$ naj predstavlja vektor z začetno točko v \mathbf{M}_1 in končno točko v \mathbf{M}_i . Oznaka θ_i naj izraža orientacijo, oznaka $l_i = \|s_i\|$ pa dolžino odseka s_i . Orientacija je vedno element intervala $[0, \pi)$. Naj N_d označuje razdaljo med najbolj oddaljenima središčnima točkama. Upoštevajoč navedene oznake, se opisnik skupine S pri

$k > 1$ zapiše kot sledeči vektor s $4k - 2$ komponentami

$$O(S) = \left(\frac{r_2^x}{N_d}, \frac{r_2^y}{N_d}, \dots, \frac{r_k^x}{N_d}, \frac{r_k^y}{N_d}, \quad \theta_1, \dots, \theta_k, \quad \frac{l_1}{N_d}, \dots, \frac{l_k}{N_d} \right). \quad (4.5)$$

Pri $k = 1$ opisnik vsebuje zgolj vrednost θ_1 . V nadaljevanju predpostavljamo $k > 1$.



Slika 4.3: Urejena skupina k robnih odsekov pri $k = 4$. Oznake \mathbf{M}_i podajajo središčne točke posameznih odsekov in vrstni red navajanja odsekov v opisniku skupine (odsek s središčno točko \mathbf{M}_i nastopa na i -tem mestu v skupini). Siva oznaka, zaznamovana s $\mathbf{p}(S)$, predstavlja težišče točk \mathbf{M}_i . Po definiciji velja $d(\mathbf{M}_1, \mathbf{p}(S)) \leq d(\mathbf{M}_i, \mathbf{p}(S))$ za vse $i \in \{1, \dots, 4\}$. Točka $\mathbf{p}(S)$ določa položaj celotne skupine, razdalja N_d pa njeno velikost.

Razdalja N_d torej služi kot normalizacijska mera opisnika, obenem pa podaja velikost skupine: $t(S) = N_d$. (Pri $k = 1$ velikost skupine podaja l_1 .) Položaj skupine je opredeljen kot težišče središčnih točk: $\mathbf{p}(S) = \frac{1}{k} \sum_{i=1}^k \mathbf{M}_i$. Iz definicije sledi, da je opisnik skupine invarianten tako na položaj kot na velikost skupine oziroma njene manifestacije v sliki. Kot argumentirajo Ferrari in sod., je opisnik s svojimi $4k - 2$ razsežnostmi zelo kompakten.

Opisnik bi bilo mogoče na enostaven način normalizirati tudi v rotacijskem smislu. Po predlogu Ferrarija in sod. bi vsako skupino zavrteli tako, da bi orientacija prvega odseka postala 0° , nakar bi bilo θ_1 možno izločiti iz opisnika. Vendar v našem kontekstu izvedba tega koraka ni smiselna, saj pri detekciji ne obravnavamo zavrtosti objektov. Naj spomnimo, da metoda Fidler in Leonardis (2007) sicer vključuje rotacijsko normalizacijo, a le v fazi učenja značilnic, kjer so razlogi za to prostorsko-časovne narave.

4.3.2.3 Primerjanje skupin

Razdalja med opisnikoma skupin S^a in S^b , ki služi kot mera za ocenjevanje različnosti med danima skupinama, je definirana kot

$$D(S^a, S^b) = w_r \sum_{i=2}^k \left\| \frac{\mathbf{r}_i^a}{N_d^a} - \frac{\mathbf{r}_i^b}{N_d^b} \right\| + w_\theta \sum_{i=1}^k \frac{1}{\pi} |\theta_i^a - \theta_i^b| + \sum_{i=1}^k \left| \log \left(\frac{l_i^a}{N_d^a} \right) - \log \left(\frac{l_i^b}{N_d^b} \right) \right|, \quad (4.6)$$

kjer prvi člen predstavlja skupno razliko med normaliziranimi relativnimi položaji ustrezajočih odsekov, drugi člen podaja skupno razliko med orientacijami odsekov, tretji člen pa meri skupno razliko med normaliziranimi dolžinami odsekov.

V implementaciji smo uteži postavili na $w_r = 2$, $w_\theta = 4$, čeprav so Ferrari in sod. razliko v relativnih položajih utežili močnejše kot razliko v orientacijah. Sprememba orientacije po naših opažanjih namreč bolj vpliva na videz skupine kot enakovredna (v smislu prispevka k skupni razdalji) sprememba relativnih položajev segmentov. Ne glede na vrednosti uteži pa se lahko zgodi, da je razdalja med opisnikoma na videz dokaj podobnih skupin razmeroma velika, lahko pa se zgodi tudi obratno.

4.3.3 Učenje značilnic

Ker je število različnih skupin robnih odsekov, pridobljenih iz učnih slik, praviloma zelo veliko, Ferrari in sod. predlagajo uporabo gručenja. Vsaka značilnica Ferrari in sod. torej ustreza gruči podobnih skupin k robnih odsekov.

Ferrari in sod. so problem iskanja gruči prevedli na problem optimalnega razbitja grafa, zgrajenega na podlagi pridobljenih skupin in njihovih medsebojnih razdalj, na množico povezanih podgrafov (klik), ki ga je mogoče približno rešiti z algoritmom *clique-partitioning* (Ferrari in sod., 2001). V naši implementaciji pa smo uporabili kar postopek aglomerativnega gručenja, ki se je obnesel tudi pri učenju značilnic Fidler in sod. (podrazdelek 4.2.4, podrazdelek *Gručenje značilnic*).

V začetku aglomerativnega postopka vsaka skupina nastopa kot samostojna gruča. Nato se v vsakem koraku združita dve najbolj podobni gruči. Postopek se izteče, ko podobnost med najbolj podobnima gručama pade pod določen prag. Podobnost med paroma gruči je opredeljena na analogen način kot v enačbi (4.4).


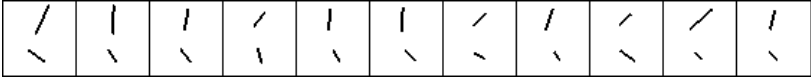

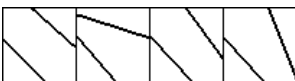

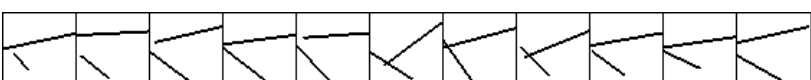




$$\text{podobnost}(C_1, C_2) = \min_{S_1 \in C_1, S_2 \in C_2} \text{podobnost}(S_1, S_2), \quad (4.7)$$

pri čemer je podobnost med skupinama definirana preprosto kot

$$\text{podobnost}(S_1, S_2) = -D(S_1, S_2). \quad (4.8)$$

V naši implementaciji je podobnostni prag, ki odloča o nadaljevanju oziroma ustavitvi gručenja, postavljen na vrednost -4 .

Tabela 4.2: Primeri značilnic in skupin $k = 2$ robnih odsekov, ki pripadajo gručam, iz katerih izvirajo posamezne značilnice.

Značilnica	Pripadajoče skupine robnih odsekov
	
	
	
	
	

Po zaključku postopka gručenja je smiselno opisnike vseh skupin k odsekov, ki pripadajo isti značilnici (gruči), strniti v enoten opisnik značilnice. Opisnik je najprikladneje definirati na tak način, da bo značilnice možno primerjati z danimi skupinami odsekov kar z uporabo (4.6). Naj $\mathcal{O} = \{O_j\}_{j=1}^n$ označuje množico n opisnikov učnih skupin, ki pripadajo dani značilnici. Poenostavimo zapis (4.5) v $O_j = (R_1^j \dots R_{2k-2}^j, \theta_1^j \dots \theta_k^j, L_1^j \dots L_k^j)$, kjer $R_{2m-3} = r_m^x/N_d$ in $R_{2m-2} = r_m^y/N_d$ za $m = 2..k$ ter $L_m = l_m/N_d$ za $m = 1..k$ (indeks j smo zaradi preglednosti izpustili). Potem je opisnik značilnice opredeljen kot

$$Z = (\bar{R}_1, \dots, \bar{R}_{2k-2}, \quad \bar{\theta}_1, \dots, \bar{\theta}_k, \quad \tilde{L}_1, \dots, \tilde{L}_k), \quad (4.9)$$

pri čemer velja $\bar{R}_m = \frac{1}{n} \sum_j R_m^j$, $\bar{\theta}_m = \frac{1}{n} \sum_j \theta_m^j$ in $\tilde{L}_m = \sqrt[n]{\prod_j L_m^j}$ (geometrijsko povprečje). Takšna definicija zagotavlja, da bo v primeru, ko je značilnica Z nastala kot gruča natanko dveh skupin, S_1 in S_2 , veljala relacija $D(S_1, Z) = D(S_2, Z) = D(S_1, S_2)/2$. Opisnik značilnice tako predstavlja nekakšno povprečje opisnikov izvornih skupin glede na definirano mero razdalje.

Tabela 4.2 predstavi nekaj primerov značilnic Ferrari in sod. Za vsako značilnico je prikazan seznam skupin $k = 2$ robnih odsekov, ki sestavljajo gručo, ki jo povzema ta značilnica.

4.3.4 Reprezentacija slik s podanimi značilnicami

V začetni fazi poteka postopek reprezentacije dane slike z dano množico značilnic \mathcal{Z} enako kot postopek pridobivanja značilnic. V sliki se z uporabo implementacije Ferrarija in sod. poiščejo skupine k približno ravnih odsekov robnih elementov. Vsaki skupini se nato priredi opisnik v skladu z zapisom (4.5). Obenem se določita položaj in velikost skupine, kot je to podano v odseku 4.3.2.2.

V naslednji stopnji postopka se vsaka dobljena skupina S reprezentira z vsemi značilnicami Z , za katere velja $D(Z, S) \leq t$ ($t = 2$ v implementaciji). Hkrati se za vsako tako značilnico določi *stopnja zanesljivosti* prireditve značilnice dani skupini. Stopnja zanesljivosti je definirana tako, da eksponentno pada z naraščajočo razliko med značilnico in skupino:

$$\text{zanesljivost}(Z, S) = \alpha^{-D(Z, S)/t}, \quad (4.10)$$

pri čemer smo v implementaciji nastavili $\alpha = 5$. Naj poudarimo, da zanesljivost ni enakovredna verjetnosti, saj v splošnem ne velja $\sum_Z \text{zanesljivost}(Z, S) = 1$. Kljub temu služi kot ustrezno nadomestilo za verjetnost $P(G | L)$ v enačbah poglavja 3.

Reprezentacija skupine S , pridobljene iz dane slike, potemtakem vsebuje naslednje podatke:

$$\begin{aligned} \text{reprezentacija}(S) &= (\mathbf{p}(S), t(S), \mathcal{R}(S)); \\ \mathcal{R}(S) &= \{(Z, \text{zanesljivost}(Z, S)) \mid Z \in \mathcal{Z}, D(Z, S) \leq t\}. \end{aligned} \quad (4.11)$$

Če v množici \mathcal{Z} ne obstaja značilnica z lastnostjo $D(Z, S) \leq t$, se skupina S izloči iz nadaljnje obravnave.

4.3.5 Uporaba značilnic Ferrari in sod. v drugih modulih sistema

Značilnice Ferrari in sod., naučene na podlagi slik dane kategorije, vstopajo v izbirni in detekcijski modul v nespremenjeni obliki.

Reprezentacija skupin k robnih odsekov z značilnicami Ferrari in sod. je nenolična, saj se isti skupini lahko priredi več značilnic. Pojem skupine k robnih odsekov popolnoma ustreza pojmu *lokalne strukture*. Verjetnost $P(G | L)$, ki nastopa v enačbah poglavja 3, se nadomesti kar z zanesljivostjo prireditve značilnice $G \in \mathcal{Z}$ skupini (lokalni strukturi) L .

Velikostna invariantnost značilnic Ferrari in sod. omogoča uporabo večvelikostnega detekcijskega postopka, opisanega v podrazdelku 3.3.2.

4.4 Povzetek poglavja

Poglavje smo namenili opisu dveh sestavov značilnic, ki smo ju vključili v implementirani sistem.

Značilnice, ki so jih zasnovali Fidler in sod., temeljijo na načelih hierarhičnosti, kompozicionalnosti in statističnosti. Hierarhična ureditev in kompozicijska zgradba

omogočata učinkovito indeksiranje kompleksnejših značilnic na osnovi enostavnejših značilnic ter prav tako učinkovito iskanje manifestacij značilnic. Zaradi fleksibilne kompozicijske predstavitve značilnic je iskanje njihovih manifestacij v slikah tudi robustno. Izhodišče za gradnjo značilnic predstavljajo elementarna Gaborjeva jedra, ki so občutljiva na usmerjene intenzitetne prehode. Tekom učenja, ki ga usmerjajo statistične lastnosti vizualne informacije v učnih slikah, se preproste in nedistinktivne značilnice sestavljajo v čedalje kompleksnejše in čedalje bolj distinktivne značilnice. Pomemben vidik učenja so tudi prijemi za brzdanje časovne in prostorske ekspanzije pri naraščanju kompleksnosti.

Značilnice, ki so jih zasnovali Ferrari in sod., pa prvenstveno temeljijo na načelu povezanosti. Izhodišče za gradnjo značilnic predstavljajo podatki o usmerjenih intenzitetnih prehodih v danih učnih slikah. Postopek gradnje nato elemente intenzitetnih prehodov po kriteriju povezanosti organizira v množico skupin približno ravnih odsekov. Značilnice nastanejo po gručenju dobljene množice skupin. Opisnik značilnice je kompakten vektor, ki preko mere razdalje omogoča enostavno primerjavo s posameznimi skupinami, pridobljenimi iz dane slike.

Eksperimentalni rezultati

Pričujoče poglavje namenjam prikazu rezultatov izbranih preizkusov predstavljenih metod. V razdelku 5.1 podamo kratek pregled slikovnih zbirk, ki so služile kot testne množice za vse opravljene preizkuse. Razdelek 5.2 namenjam podrobnejšemu opisu načina izvajanja preizkusov. V naslednjih štirih razdelkih navajamo in komentiramo rezultate glavnega eksperimentalnega niza, v okviru katerega smo preizkusili vse štiri kombinacije sestavov značilnic in izbirnih metod:

- kombinacija značilnic Fidler in sod. in filterske izbirne metode (razdelek 5.3);
- kombinacija značilnic Fidler in sod. in notranjeoptimizacijske izbirne metode (razdelek 5.4);
- kombinacija značilnic Ferrari in sod. in filterske izbirne metode (razdelek 5.5);
- kombinacija značilnic Ferrari in sod. in notranjeoptimizacijske izbirne metode (razdelek 5.6).

Razdelek 5.7 povzema poglavje in najpomembnejše izsledke.

5.1 Testne slikovne množice

Za izvajanje preizkusov smo uporabili sledeče slikovne zbirke:

- Avtomobili UIUC (Agarwal in sod., 2004)
(<http://l2r.cs.uiuc.edu/~cogcomp/Data/Car>).
- Konji INRIA (Ferrari in sod., 2006a)
(<http://pascal.inrialpes.fr/data/horses>).
- Logotipi blagovne znamke Apple iz zbirke ETHZ (Ferrari in sod., 2006a)
(<http://www.vision.ee.ethz.ch/~ferrari>).
- Steklenice ETHZ (Ferrari in sod., 2006a)
(<http://www.vision.ee.ethz.ch/~ferrari>).

- Skodelice ETHZ (Ferrari in sod., 2006a) (<http://www.vision.ee.ethz.ch/~ferrari>).

Vsaka slikovna zbirka opredeljuje testno množico za ustrezno vizualno kategorijo. Posledično vsaka zbirka definira samostojen detekcijski problem.

Navedimo nekaj lastnosti testnih zbirk:

- Za vse testne množice velja, da vsaka slika vsebuje najmanj en objekt iz ustrezne kategorije. Vendar pa se detekcijski postopek, kot smo omenili v podrazdelku 1.1.1, ne sme posluževati nikakršnih predpostavk o številu objektov na dani vhodni sliki, temveč mora upoštevati, da je to število lahko poljubno (vštevši nič). Detekcijski pristop, opisan v poglavju 3, se tega pravila drži.
- Vsi objekti v vseh zbirkah so upodobljeni s stranskega pogleda in v „naravni“ zavrtosti okrog osi, pravokotne na slikovno ravnino. Vendar pa so pri posameznih slikah možna manjša odstopanja v pogledu in zavrtosti.
- Objekti zbirke avtomobilov UIUC so prikazani tako z leve kot z desne strani (avtomobil je upodobljen bodisi tako, da slika prikazuje njegovo desno stran, ali pa tako, da slika prikazuje njegovo levo stran). Pri vseh ostalih zbirkah so objekti upodobljeni le z ene strani. Pri zbirki skodelic, na primer, se kot objekti obravnavajo le podobe skodelic z ročajem na desni strani. Čeprav zbirka vsebuje tudi nekaj podob skodelic z ročajem na levi strani, se takšne podobe obravnavajo kot del ozadja. Morebitna detekcija skodelice z ročajem na levi bi se torej obravnavala kot lažni alarm. Enako velja za zbirke konj, steklenic in logotipov Apple.
- Objekti so v vseh testnih zbirkah upodobljeni pri različnih velikostih. Največji velikostni razpon objektov izkazuje množica logotipov Apple, najmanjši razpon pa množica avtomobilov.
- Največjo raznolikost izgleda objektov najbrž izkazuje množica konj. Zbirka vsebuje konje različnih pasem, barv in upodobljenih razporeditev okončin. Najmanjšo raznolikost izgleda izkazuje množica logotipov Apple. Kljub razmeroma velikem razponu velikosti upodobljenih objektov je množica logotipov Apple verjetno najenostavnejša med vsemi petimi testnimi zbirkami.

Nekateri podatki o testnih množicah so zbrani v tabeli 5.1. Pojasnimo podatke v tabeli na primeru zbirke avtomobilov. Testna množica sestoji iz 108 slik, ki skupno prikazujejo 139 objektov. Ker je širina podob avtomobilov s stranskega pogleda v povprečju večja od višine, smo za referenčno mero za določanje velikosti objektov izbrali širino omejevalnega okvirja. Omejevalni okvir podobe najmanjšega (v smislu referenčne mere) objekta v testni množici je širok 199 slikovnih elementov, okvir največjega objekta pa 516 slikovnih elementov. Povprečna širina

Tabela 5.1: Izbrani podatki o testnih slikovnih množicah.

Testna zbirka	SL ^a	OBJ ^b	REF ^c	Vel. razpon ^d	Povp. vel. ^e	Povp. razm. ^f
Avtomobili	108	139	širina	199 do 516	$339 \pm 25\%$	$3,36 \pm 6\%$
Konji	170	180	širina	74 do 685	$292 \pm 43\%$	$1,18 \pm 18\%$
Logotipi Apple	40	44	višina	59 do 429	$172 \pm 52\%$	$1,22 \pm 11\%$
Steklenice	48	55	višina	131 do 530	$345 \pm 27\%$	$3,40 \pm 13\%$
Skodelice	48	66	širina	79 do 410	$201 \pm 43\%$	$1,02 \pm 16\%$

^aŠtevilo slik v zbirki.^bSkupno število objektov v zbirki.^cReferenčna mera za opredeljevanje velikosti objektov.^dMinimum in maksimum referenčne mere.^ePovprečna vrednost referenčne mere \pm standardni odklon.^fPovprečno razmerje širina : višina (REF = širina) oz. višina : širina (REF = višina).

omejevalnih okvirjev objektov v zbirki znaša 339 slikovnih elementov s standardnim odklonom 84 slikovnih elementov ($\approx 25\%$). Povprečno razmerje med širino in višino omejevalnega okvirja znaša 3,36 s standardnim odklonom 0,18 ($\approx 6\%$).

Slika 5.1 prikazuje primere slik iz vseh petih testnih množic. Slike so prikazane v enaki višini iz estetskih razlogov. Dejansko se razsežnosti slik med seboj precej razlikujejo. Enako velja tudi za vse nadaljnje prikaze slik v tem poglavju.

V fazi predobdelave smo s pomočjo Matlabove funkcije `rgb2gray` vse barvne testne slike pretvorili v sivinske. Ker so podobe avtomobilov na slikah zbirke UIUC zelo majhne, smo vsako sliko iz te zbirke povečali za faktor 2,5 v vsaki od obeh koordinatnih smeri. Pri tem smo uporabili bikubično interpolacijo, ki jo izvaja Matlabova funkcija `imresize`. V testni fazi drugih predobdelovalnih operacij nismo izvedli.

5.2 Podrobnosti izvajanja preizkusov

5.2.1 Shema izvajanja preizkusov glavnega niza

Za vsako kombinacijo izbirne metode, sestava značilnic in testne množice (oz. vizualne kategorije, ki ustreza testni množici) smo izvedli niz neodvisnih preizkusov. Vsak preizkus glavnega eksperimentalnega niza je potekal na sledeč način:

1. Postopek za pridobivanje značilnic ustvari izhodiščno množico značilnic (\mathcal{Z}) na podlagi učne slikovne množice \mathcal{I}_z za dano vizualno kategorijo. Pridobljena množica značilnic se uporabi za reprezentacijo vseh učnih in testnih slik dane kategorije.
2. Na podlagi učne slikovne množice \mathcal{I}_d se zgradi detekcijski model za vsako značilnico iz \mathcal{Z} .
3. Z uporabo postopka za izbiranje značilnic in učne slikovne množice \mathcal{I}_i se iz množice \mathcal{Z} izbere nabor T_{\max} značilnic, kjer T_{\max} predstavlja največje število



(a) Avtomobili



(b) Konji (slika se nadaljuje na naslednji strani)



(c) Logotipi Apple



(d) Steklenice



(e) Skodelice

Slika 5.1: Primeri slik iz vseh uporabljenih testnih zbirk.

izbranih značilnic, ki ga še želimo obravnavati. Ker vse predstavljene izbirne metode vračajo zaporedja izbranih značilnic, ki se s povečevanjem ciljnega števila izbranih značilnic (T) zgolj podaljšujejo, zaporedje T_{\max} značilnic že vsebuje izbirne množice za vse $T \in \{1, \dots, T_{\max}\}$.

4. V testni fazi se za vsako vrednost parametra T izvede detekcijski postopek na celotni testni množici. Za reprezentiranje testnih slik se uporabi celotna množica \mathcal{Z} , vendar pa sme detekcijski postopek pri podanem T graditi detekcijske hipoteze izključno na podlagi manifestacij izbranih T značilnic v testnih slikah. Za vsak T in vsako testno množico se izračuna krivulja RPC, ki se ovrednoti z merama F_{\max} in EER. Krivulja odvisnosti F_{\max} od števila izbranih značilnic T se obravnava kot karakteristična krivulja preizkusa. V nadaljevanju bomo prikazali karakteristično krivuljo za vsako kombinacijo izbirne metode, sestava značilnic in testne množice.

Ker učna množica \mathcal{I}_z v vseh primerih sovпада z množico \mathcal{I}_d , bomo ti dve množici obravnavali kot eno samo množico, ki jo bomo označevali kot \mathcal{I}_{zd} .

V nadaljevanju bomo podali nekaj podrobnosti izvajanja preizkusov. Podrazdelek 5.2.2 se nanaša na kombinacijo filterske izbirne metode in značilnic Fidler in sod., podrazdelek 5.2.3 pa na kombinacijo filterske izbirne metode in značilnic Ferrari in sod. Podrazdelek 5.2.4 opiše nekaj izvedbenih podrobnosti za primer uporabe notranjeoptimizacijske izbirne metode. Podrazdelek 5.2.5 navaja vrednosti nekaterih pomembnih parametrov sistema.

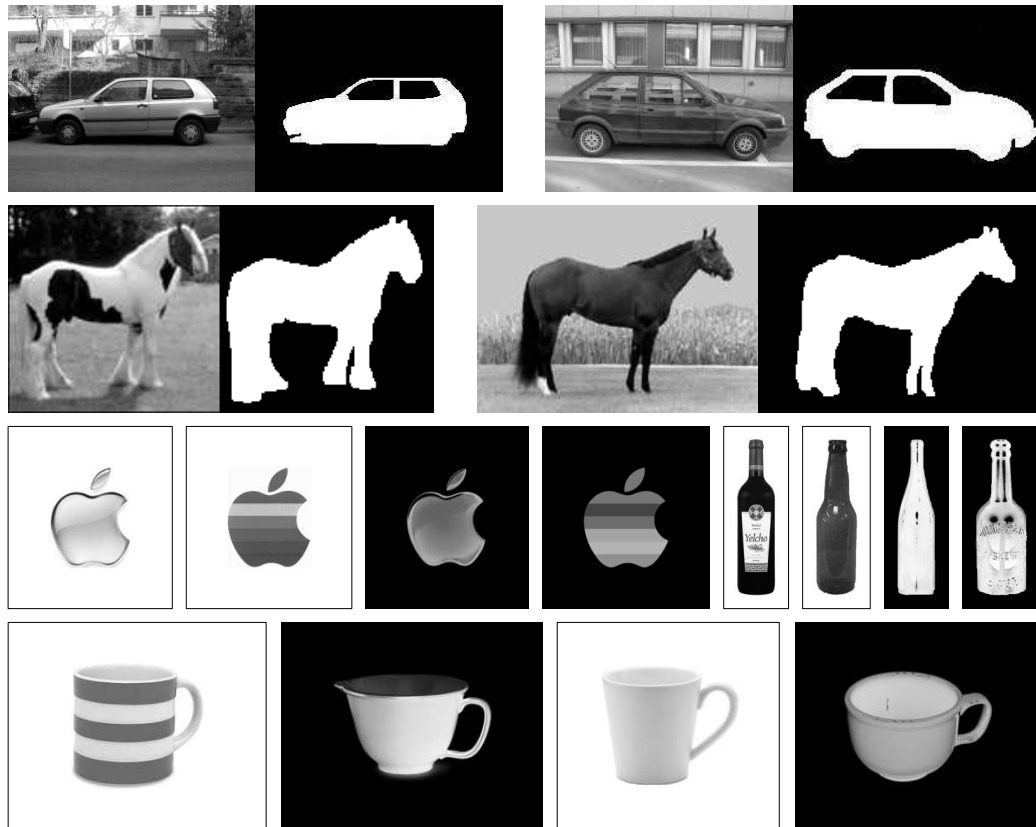
5.2.2 Filterska izbira značilnic Fidler in sod.

V tem podrazdelku navajamo nekaj podrobnosti izvajanja preizkusov kombinacije filterske izbirne metode in značilnic Fidler in sod. Odseka 5.2.2.1 in 5.2.2.2 se nanašata na učno fazo, odsek 5.2.2.3 pa podaja izvedbene podrobnosti testne faze.

5.2.2.1 Množice \mathcal{I}_{zd}

Nekaj primerov slik iz učnih množic \mathcal{I}_{zd} je prikazanih na sliki 5.2. Nekateri podatki o učnih množicah so nanizani v tabeli 5.2. Pri vsaki množici sta navedena spletni vir in sklic na ustrezno objavo. Učne množice za logotipe Apple, steklenice in skodelice so bile pridobljene z iskanjem po spletu s pomočjo iskalnika Google.

Kot smo omenili v podrazdelku 3.2.1, se v fazi predobdelave učne slike za dano kategorijo velikostno normalizirajo tako, da omejevalni okvirji vseh objektov v učni množici postanejo bodisi enako široki bodisi enako visoki. Pri učni množici kategorije avtomobilov, na primer, so po normalizaciji omejevalni okvirji vseh podob avtomobilov široki 220 slikovnih elementov. Postopek velikostne normalizacije seveda spremeni višino in širino dane slike za enak faktor, saj bi se sicer oblika objekta popačila.



Slika 5.2: Primeri učnih slik iz množice \mathcal{I}_{zd} za posamezne vizualne kategorije za kombinacijo filtrske izbirne metode in značilnic Fidler in sod. Slike avtomobilov in konj so prikazane na poljubnih ozadjih, zato vsi moduli učne faze uporabljajo priložene segmentacijske maske. V prvih dveh vrstah so prikazane skupno štiri slike avtomobilov in konj ter ustrezajoče segmentacijske maske. Pri ostalih preizkušanih kategorijah sta postopek za pridobivanje značilnic in postopek za učenje detekcijskega modela uporabljala slike, ki prikazujejo objekte na čistem enobarvnem ozadju. Segmentacijske maske zato niso bile potrebne.

Tabela 5.2: Nekateri podatki o učnih množicah \mathcal{I}_{zd} za posamezne vizualne kategorije za kombinacijo filtrske izbirne metode in značilnic Fidler in sod. V zadnjem stolpcu sta za vsako učno množico navedeni razsežnost, vzdolž katere se izvaja velikostna normalizacija, in ciljna normalizirana velikost omejevalnih okvirjev podob objektov.

Kategorija	Vir	Št. slik	Normalizacija
Avtomobili	Leibe in sod. (2004) ^a	100	širina, 220
Konji	Borenstein in Ullman (2002) ^b	328	širina, 140
Logotipi Apple	svetovni splet	54	višina, 120
Steklenice	svetovni splet	66	višina, 220
Skodelice	svetovni splet	184	širina, 150

^a<http://www.pascal-network.org/challenges/VOC/databases.html>

^b<http://www.msri.org/people/members/eranb>



Slika 5.3: Po dva primera pozitivnih in negativnih slik učne zbirke UIUC.

Značilnice Fidler in sod. so občutljive na polarnost intenzitetnih prehodov: prehod „levo temno, desno svetlo“ sproži drugačne odzive kot prehod „levo svetlo, desno temno“. Zato je smiselno, če učna množica vsebuje slike, ki prikazujejo objekte na svetlih ozadjih, in slike, ki prikazujejo objekte na temnih ozadjih. Pri množicah za logotipe Apple, steklenice in skodelice smo to priporočilo uveljavili enostavno z dodajanjem sivinsko invertiranih različic vseh slik v množico \mathcal{I}_{zd} . Množica \mathcal{I}_{zd} za logotipe Apple, na primer, sestoji iz 27 slik logotipov na belem ozadju in 27 invertiranih različic teh slik. Pri učnih množicah za avtomobile in konje tovrsten korak ni bil potreben, saj so podobe objektov v slikovni zbirki obdane z zelo raznolikimi ozadji, zaradi česar se na robovih objektov pojavljajo intenzitetni prehodi različnih polarnosti. Poleg tega so izdelovalci učnih množic za avtomobile in konje za vsako sliko ustvarili segmentacijsko masko, ki omogoča ločitev manifestacij značilnic, ki pripadajo objektu, od manifestacij, ki pripadajo ozadju.

5.2.2.2 Množice \mathcal{I}_i

Filtrska metoda za izbiro značilnic, ki smo jo predstavili v razdelku 2.3, predpostavlja, da množica učnih slik \mathcal{I}_i , na podlagi katere se izvaja izbirni postopek, sestoji iz množice pozitivnih slik (\mathcal{I}_+) in iz množice negativnih slik (\mathcal{I}_-). Pri kategoriji avtomobilov smo kot množico \mathcal{I}_i obravnavali učno zbirko UIUC (Agarwal in sod., 2004), ki jo sestavlja 550 pozitivnih in 500 negativnih slik. Po dva primera pozitivnih in negativnih slik učne zbirke UIUC sta prikazana na sliki 5.3.



Slika 5.4: Primeri slik iz zbirke GRAZ-01 (prve tri slike v prvi vrsti), iz zbirke LPP-34 (zadnja slika v prvi vrsti in prvi dve slike v drugi vrsti) in iz raznih spletnih virov (zadnja slika v drugi vrsti).

Tabela 5.3: Število pozitivnih in negativnih slik v posameznih množicah \mathcal{I}_i za kombinacijo filtrske izbirne metode in značilnic Fidler in sod.

Kategorija	$ \mathcal{I}_+ $	$ \mathcal{I}_- $
Avtomobili	550	500
Konji	328	182
Logotipi Apple	54	420
Steklenice	66	366
Skodelice	184	224

Pri vseh ostalih kategorijah pozitivna podmnožica množice \mathcal{I}_i (torej \mathcal{I}_+) sovpada z učno množico \mathcal{I}_{zd} . Gradivo za negativne podmnožice \mathcal{I}_i (torej \mathcal{I}_-) smo črpali iz množic \mathcal{I}_+ za druge kategorije, iz slikovne zbirke GRAZ-01 (Opelt in sod., 2006a), iz zbirke LPP-34 (Omerčević in sod., 2007) in iz slik, pridobljenih s pomočjo spletnega iskalnika Google. Slika 5.4 prikazuje nekaj primerov slik iz omenjenih virov. Tabela 5.3 navaja podatke o številu slik v posameznih množicah \mathcal{I}_i .

Po opredelitvi učnih slikovnih množic se izvede postopek izbire značilnic. Najprej se iz slik učne množice \mathcal{I}_{zd} pridobi izhodiščni nabor značilnic. Za vsako pridobljeno značilnico se nato zgradi detekcijski model na osnovi iste učne množice. Na osnovi naučenih detekcijskih modelov in slik \mathcal{I}_i se izvrši dvostopenjski postopek izbire značilnic. V prvi stopnji se za vsako kombinacijo učne slike $I \in \mathcal{I}_i$ in pridobljene značilnice Z izračuna vrednost funkcije $S(Z, I)$, v drugi stopnji pa postopek AdaBoost proizvede zaporedje izbranih značilnic $\langle Z^{[1]}, Z^{[2]}, \dots, Z^{[T_{\max}]} \rangle$.

Za izračun funkcije $S(Z, I)$ se izvrši poenostavljena različica testnega postopka. Ker so objekti v množici \mathcal{I}_+ normalizirani na enako velikost kot objekti v množici \mathcal{I}_{zd} , smo pri izračunu funkcije S uporabili kar enovelikostni detekcijski postopek.

Poleg tega se pri funkciji S upošteva le najmočnejša detekcijska hipoteza, zato je uporaba postopka MDL odveč.

5.2.2.3 Testna faza

V testni fazi se za vsak $T \in \{1, \dots, T_{\max}\}$ izvede detekcijski postopek na vsaki testni množici. Kot smo že omenili, značilnice Fidler in sod. uporabljamo v kombinaciji z detekcijskim postopkom iz razdelka 3.2. Ker detekcijski postopek ne pozna velikosti objektov v testnih slikah, mora vsako testno sliko obravnavati v okviru velikostne piramide. Razmak med sosednjima plastema piramide smo povsod nastavili na $\sqrt[4]{2}$, kar ustreza štirim plastem na oktavo ($R = 4$). Število plasti piramide pa je odvisno od kategorije oz. testne množice. Pri avtomobilih, kjer je pričakovani velikostni razpon testnih objektov razmeroma majhen, smo število piramidnih plasti nastavili na 7. Za detekcijo steklenic in skodelic smo uporabili 11, za detekcijo konj in logotipov Apple pa 12 plasti.

Pri avtomobilih spodnja plast piramide ustreza pričakovani velikosti testnega objekta v izmeri $2^{-1/4}$ učne velikosti, zgornja plast pa pričakovani velikosti v izmeri $2^{+5/4}$ učne velikosti. Spodnja plast torej vsebuje $2^{1/4}$ -krat povečano različico testne slike, velikost različice na zgornji plasti pa je enaka $2^{-5/4}$ -kratniku velikosti izvirne testne slike. Ker so pozitivne učne slike za kategorijo avtomobilov normalizirane tako, da širina prikazanih objektov oz. njihovih omejevalnih okvirjev znaša 220 slikovnih elementov, spodnja plast piramide ustreza pričakovani velikosti (širini) testnega avtomobila okrog $220 \times 2^{-1/4} \approx 185$ slikovnih elementov, zgornja plast pa ustreza velikosti okrog $220 \times 2^{+5/4} \approx 523$ slikovnih elementov. Detekcijski postopek je tako zmožen odkriti avtomobile z velikostmi (širinami) od 185 do 523 slikovnih elementov. Zaradi robustnosti značilnic in detekcijskega postopka na majhna velikostna odstopanja pa je ta interval v resnici še nekoliko širši.

V tabeli 5.4 so zbrani podatki o velikostnih piramidah za vse testne zbirke. Seveda se vse slike iz dane testne množice obravnavajo na enak način, torej v okviru velikostne piramide z istimi parametri.

5.2.3 Filtrska izbira značilnic Ferrari in sod.

Značilnice Ferrari in sod. so zasnovane in opisane na velikostno invarianten način, zato bi načeloma bilo možno obravnavati testne slike le pri njihovi izvirni velikosti. Za dosego večje robustnosti pa je, kot smo navedli v odseku 3.3.2.4, smiselno obravnavati učne ali testne slike (ali oboje) pri več različnih velikostih.

V vseh učnih množicah \mathcal{I}_{zd} nastopa vsaka slika v petih velikostnih različicah, in sicer v $2^{-4/4}$ -, $2^{-2/4}$ -, $2^{0/4}$ -, $2^{+2/4}$ -in $2^{+4/4}$ -kratnikih izhodiščne učne velikosti za vizualno kategorijo. Podatki o številu slik v množicah \mathcal{I}_{zd} in o izhodiščnih učnih velikostih so zbrani v tabeli 5.5. V testni fazi se vsaka slika obravnava v njeni izvirni velikosti ter v $2^{-2/4}$ - in $2^{+2/4}$ -kratnikih njene izvirne velikosti.

Tabela 5.4: Razpon velikostne piramide za posamezne testne zbirke pri izvajanju testne faze za kombinacijo filtrske izbirne metode in značilnic Fidler in sod. Navedene so le oznake eksponentov. Opisnik „a do b“, na primer, predstavlja dejstvo, da pričakovana velikost objekta na spodnji plasti piramide ustreza $2^{a/4}$ -kratniku učne velikosti, pričakovana velikost na zgornji plasti pa $2^{b/4}$ -kratniku učne velikosti. Število plasti piramide znaša $b - a + 1$.

Testna zbirka	Razpon velikostne piramide
Avtomobili	−1 do +5
Konji	−1 do +10
Logotipi Apple	−3 do +8
Steklenice	−3 do +7
Skodelice	−3 do +7

Tabela 5.5: Število slik in izhodiščna učna velikost za učne množice \mathcal{I}_{zd} za kombinacijo filtrske izbirne metode in značilnic Ferrari in sod. Vsaka učna slika je predstavljena v petih različnih velikostih (izhodiščna velikost ter njeni 2-, $\sqrt{2}$ -, $(1/\sqrt{2})$ - in $(1/2)$ -kratniki), kar je nakazano z oznako „ $\times 5$ “.

Kategorija	$ \mathcal{I}_{zd} $	Izhodiščna velikost
Avtomobili	58×5	326 (širina)
Konji	40×5	260 (širina)
Logotipi Apple	27×5	170 (višina)
Steklenice	33×5	310 (višina)
Skodelice	34×5	210 (širina)



Slika 5.5: *Primeri avtomobilov iz Yahoojeve slikovne zbirke.*

Viri za učne množice za kategorije konj, logotipov Apple, steklenic in skodelic so enaki odgovarjajočim virom pri značilnicah Fidler in sod. Vir za učno množico za kategorijo avtomobilov pa je Yahoojeva slikovna zbirka¹. Primeri učnih avtomobilov sta prikazana na sliki 5.5. Značilnice Fidler in sod. v nasprotju z značilnicami Ferrari in sod. niso občutljive na polarnost intenzitetnega prehoda, zato sivinsko invertiranje ni potrebno. Vse učne slike so prikazane na belem ozadju, slike iz zbirke podob konj (Borenstein in Ullman, 2002) pa so opremljene s segmentacijskimi maskami.

Vse učne množice \mathcal{I}_i so sestavljene iz 100 pozitivnih in 100 negativnih slik. Pri vsaki kategoriji je nabor \mathcal{I}_+ pridobljen kot naključna podmnožica ustrežajoče množice \mathcal{I}_{zd} , pri čemer se velikostne razlike iste slike upoštevajo kot različne slike. Nabori \mathcal{I}_- sestojijo iz naključno velikih kvadratnih zaplat slik iz zbirk GRAZ-01, LPP-34 in zbirke, pridobljene s pomočjo iskalnika Google.

Za detekcijo objektov na podlagi značilnic Ferrari in sod. se uporablja detekcijski postopek iz podrazdelka 3.3.2, v katerega je večvelikostnost vgrajena že po zasnovi. Isti detekcijski postopek se uporablja tudi pri računanju funkcije S v fazi izbiranja značilnic. V slednjem primeru se dušilni postopek MDL ne izvaja, saj funkcija S temelji le na moči najmočnejše hipoteze za dano sliko.

5.2.4 Notranjeoptimizacijska izbira

Izbirna metoda na osnovi notranje optimizacije poganja detekcijski postopek za različne kandidatne množice značilnic. Ker je tovrstna metoda celo pri požrešnem preiskovalnem postopku časovno zelo zahtevna, je nujno privzeti določene predpostavke in poenostavitve. Pri vseh kategorijah množica \mathcal{I}_i sestoji zgolj iz 20 slik. Deset slik vsebuje le različna ozadja, preostalih deset slik pa prikazuje podobe objektov na ozadjih. Slike za ozadja so pridobljene iz zbirk GRAZ-01 in LPP-34 ter iz zbirke naključnih slik s spleta. Pri vseh kategorijah je skupno število prikazanih objektov v desetih „pozitivnih“ slikah enako 20. Primeri „pozitivnih“ slik so prikazani na sliki 5.6, primeri „negativnih“ slik pa so prikazani na sliki 5.4. „Pozitivne“ slike smo ustvarili preprosto z lepljenjem objektov iz množic \mathcal{I}_{zd} na različna ozadja. (Naj spomnimo, da pri notranjeoptimizacijski izbiri množica \mathcal{I}_i ne sestoji iz pozitivnih in negativnih slik v takšnem smislu kot pri filterski izbiri. Slikovna množica \mathcal{I}_i bi lahko vsebovala zgolj slike, ki prikazujejo objekte. Besedni

¹<http://autos.yahoo.com>



Slika 5.6: Primeri „pozitivnih“ učnih slik iz množic \mathcal{I}_i za primer notranjeoptimizacijske izbirne metode. Slike v zgornji vrsti nastopajo v posameznih množicah \mathcal{I}_i za sestav značilnic Fidler in sod. Spodnja vrsta predstavlja primere „pozitivnih“ slik iz množic \mathcal{I}_i za sestav značilnic Ferrari in sod.

zvezi „pozitivne“ slike in „negativne“ slike služita le enostavnejšemu izražanju.)

V primeru značilnic Fidler in sod. izbirni algoritem izvaja enovelikostni detekcijski postopek, zato morajo biti objekti, ki nastopajo na „pozitivnih“ slikah množice \mathcal{I}_i , prikazani v normalizirani učni velikosti (tabela 5.2, zadnji stolpec). V primeru značilnic Ferrari in sod. pa se izvaja večvelikostni detekcijski postopek, zato je smiselno, če slike vsebujejo objekte različnih velikosti.

Mera za ocenjevanje dane kandidatne množice značilnic oz. danega vozlišča v prostoru stanj je definirana kot $(F_{\max} + \text{EER})/2$ za krivuljo RPC, ki izraža rezultat izvedbe detekcijskega postopka, omejenega na uporabo dane kandidatne množice značilnic, na množici \mathcal{I}_i .

V prvi iteraciji požrešnega izbirnega postopka se detekcijski postopek izvrši za vse posamične značilnice. V drugo iteracijo pa vstopijo le značilnice, z uporabo katerih je detekcijski postopek na množici \mathcal{I}_i dosegel $\text{EER} > 0$ (značilnice Fidler in sod.) oziroma $F_{\max} > 0$ (značilnice Ferrari in sod.). Čeprav obstaja možnost, da se na tak način iz nadaljnje obravnave izloči značilnica, ki bi se utegnila obnesti v kombinaciji z drugimi značilnicami, je spričo velike časovne zahtevnosti takšna poenostavitvena heuristika skoraj nujna. Z izločitvijo značilnic, ki posamično vodijo do slabih detekcijskih rezultatov, se prostor stanj močno skrči. Ker je pri vseh kategorijah izhodiščna množica značilnic Fidler in sod. večja od ustrezne izhodiščne množice značilnic Ferrari in sod., smo pri sestavu Fidler in sod. uporabili agresivnejši izločitveni pogoj. F_{\max} je namreč vselej večji od EER pri isti krivulji RPC, zato je teže doseči $\text{EER} > 0$ kot $F_{\max} > 0$.

Nadaljnje prihranke pri porabi računskega časa smo omogočili z zmanjšanjem ločljivosti hipoteznega prostora v obeh prostorskih razsežnostih, poleg tega pa smo se odpovedali postopku MDL. Vse našete poenostavitve so nujne za zagotovitev

praktične izvedljivosti notranjeoptimizacijskega izbirnega postopka. Seveda pa se okrnjena različica detekcijskega postopka izvaja le v fazi izbiranja značilnic. V testni fazi se uporablja neokrnjen detekcijski postopek, ki vključuje tudi postopek MDL.

Požrešni postopek, prikazan kot algoritem 2.2, se ustavi, ko detekcijske točnosti ni več mogoče izboljšati. Pri izvajanju preizkusov pa smo uporabili za odtenek drugačen pristop: požrešni postopek se v vsaki iteraciji pomakne v najboljše ocenjenega naslednika trenutnega stanja, ustavi pa se šele, ko število izbranih značilnic doseže ciljno vrednost T_{\max} . Izkaže se namreč, da je zaradi majhnega števila slik že po nekaj iteracijah mogoče doseči „strop“, po katerem se ocena $(F_{\max} + \text{EER})/2$ ne izboljšuje več.

5.2.5 Vrednosti nekaterih parametrov

Vse preizkuse s sestavom značilnic Fidler in sod. smo izvedli z uporabo značilnic ravni 3. Zaradi kompozicijske zgradbe značilnic pa množica značilnic ravni 3 lahko posredno vključuje tudi značilnice ravni 2 in 1. V procesu statističnega učenja se namreč lahko zgodi, da neka značilnica ravni 1 nastopa kot edina komponenta neke značilnice ravni 2, ta pa nastopa kot edina komponenta neke značilnice ravni 3.

Pri preizkusih, ki so jih izvedli Ferrari in sod. (2006a), so se najbolj obnesle značilnice, nastale na podlagi skupin dveh robnih odsekov. Vrednost parametra k pri značilnicah Ferrari in sod. smo zato nastavili na 2.

Vrednosti nekaterih drugih parametrov so bile določene s preizkušanjem na neki učni množici. V naši implementaciji sestava Ferrari in sod. igra pomembno vlogo prag t , ki določa največjo razdaljo med značilnico Z in iz slike pridobljeno skupino robnih odsekov S , pri kateri se skupina S še obravnava kot manifestacija značilnice Z . Če je prag postavljen prenizko, reprezentacija slik z značilnicami Ferrari in sod. ni dovolj robustna na odstopanja od strukture značilnice. Če je prag postavljen previsoko, se lahko zgodi, da se skupini v fazi reprezentacije priredi (tudi) značilnica, ki ji vizualno ni dovolj podobna. Kot smo omenili v podrazdelku 4.3.3, smo prag t nastavili na 2, saj je ta vrednost v povprečju omogočala najboljše detekcijske rezultate na učnih množicah za posamezne kategorije.

S preizkušanjem na učni množici smo določili tudi vrednosti parametrov Gaussovega glajenja hipoteznega prostora v fazi detekcije objektov.

Vrednosti pomembnejših parametrov metode so povzete v tabeli 5.6.

5.3 Rezultati filtrske izbire značilnic Fidler in sod.

5.3.1 Rezultati preizkusov

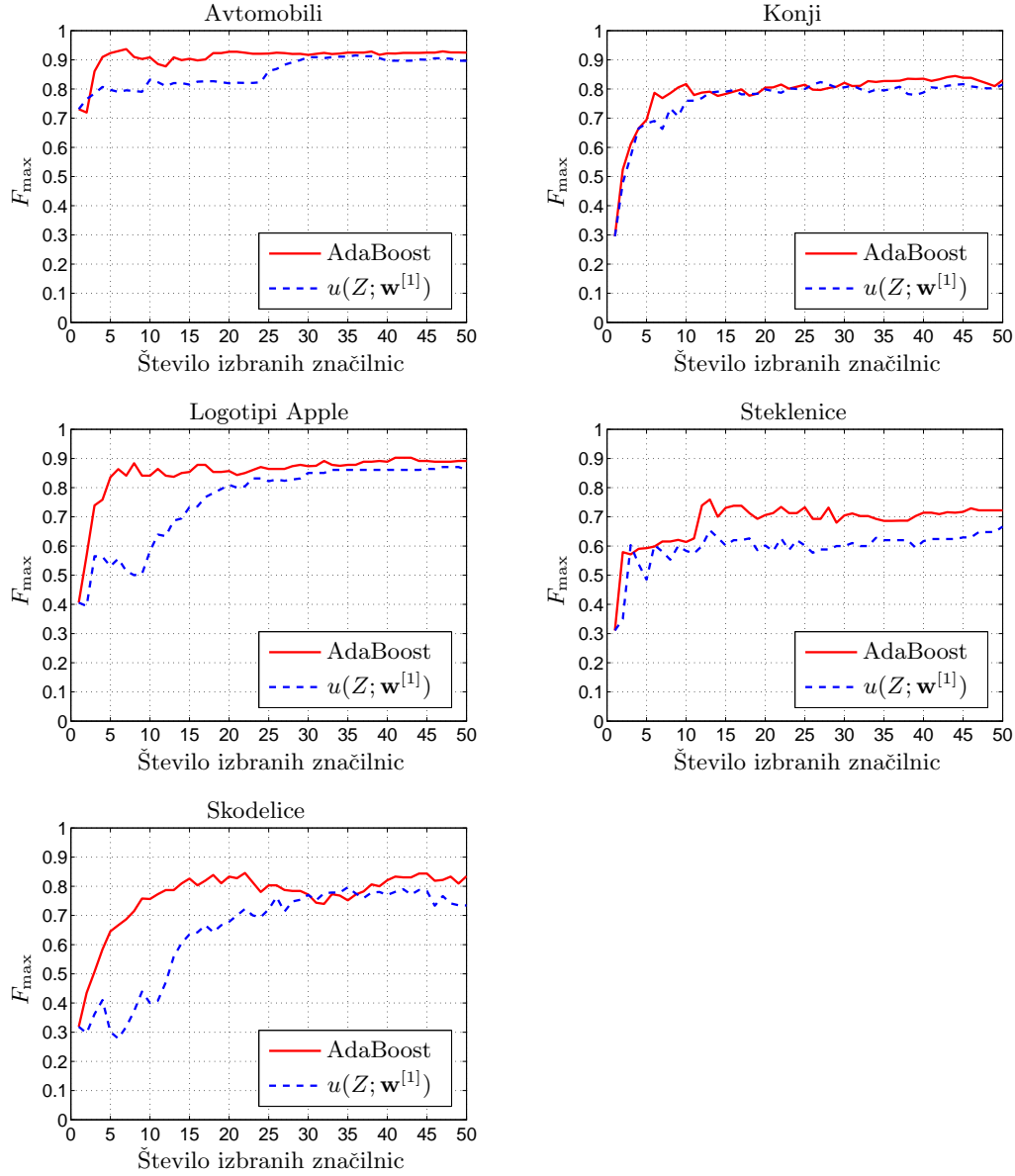
Slika 5.7 prikazuje odvisnost karakteristike F_{\max} krivulje RPC od števila izbranih značilnic (T) pri kombinaciji filtrske izbirne metode in značilnic Fidler in sod. Na sliki primerjamo detekcijske sposobnosti zaporedja prvih 50 izbranih značilnic,

Tabela 5.6: Vrednosti izbranih parametrov metode.

Detekcijski modul		
Parameter	Vrednost	Opombe
σ	2	Gaussov filter pri enovelikostni detekciji
R	4	število plasti na oktavo
σ_r	1	3-razsežni filter pri večvelikostni detekciji
$\sigma_x(r)$	$3\sqrt{2^{r/R}}$	3-razsežni filter pri večvelikostni detekciji
$\sigma_y(r)$	$3\sqrt{2^{r/R}}$	3-razsežni filter pri večvelikostni detekciji
Značilnice Fidler in sod.		
Parameter	Vrednost	Opombe
λ	1	razmerje polosi Gaborjevega jedra Z_1^1
σ	2,5	ravninski razteg Gaborjevega jedra Z_1^1
F	2/35	frekvenca Gaborjevega jedra Z_1^1
#orientacij	8	število orientacij
prag pri gručenju	0,85	(odsek 4.2.4.3)
n	3	hierarhična raven uporabljenih značilnic
Značilnice Ferrari in sod.		
Parameter	Vrednost	Opombe
w_r	2	utež pri računanju razdalje (4.6)
w_θ	4	utež pri računanju razdalje (4.6)
k	2	število robnih odsekov na skupino
prag pri gručenju	-4	(podrazdelek 4.3.3)
t	2	parameter v (4.10)
α	5	parameter v (4.10)

kot ga je vrnil postopek AdaBoost (algoritem 2.1), in zaporedja 50 značilnic, najboljših po statičnih ocenah $u(Z; \mathbf{w}^{[1]})$. Naj spomnimo, da smo ocenjevalno funkcijo u vpeli v postopek AdaBoost zato, da bi z upoštevanjem odvisnosti med značilnicami zmanjšali redundanco v nastalih množicah izbranih značilnic. Po pričakovanju se postopek AdaBoost izkaže bolje kot izbor značilnic po padajočih ocenah $u(Z; \mathbf{w}^{[1]})$. V nadaljnjem besedilu zato praviloma navajamo in obravnavamo le rezultate izbiranja po postopku AdaBoost, razen v primerih, ko je izrecno navedeno drugače.

Detekcijska točnost vsaj na začetku razmeroma enakomerno narašča s številom značilnic, na katere se sme opreti detekcijski postopek. Vendar pa to še zdaleč ni pravilo, saj lahko ena sama neustrezno izbrana značilnica močno poslabša detekcijske rezultate in pokvari krivuljo $F_{\max}(T)$. Poleg tega so testne množice povsem ločene od učnih množic, zato ni nujno, da se značilnice, ki so bile izbrane na podlagi učnih množic, obnesejo tudi na testnih množicah. Spričo navedenih dejstev krivulje niso monotono naraščajoče.



Slika 5.7: Odvisnost F_{\max} krivulj RPC od števila izbranih značilnic za vsako posamezno vizualno kategorijo. Polne rdeče krivulje ponazarjajo detekcijske rezultate pri značilnicah, izbranih s postopkom AdaBoost, črtkane modre krivulje pa rezultate pri značilnicah, izbranih po padajočih vrednostih $u(Z; \mathbf{w}^{[1]})$. Krivulje se nanašajo na kombinacijo sestava značilnic Fidler in sod. in filtrske izbirne metode.

Tabela 5.7: *Detekcijski rezultati za slikovno zbirko avtomobilov UIUC, kot jih navajajo nekateri drugi avtorji.*

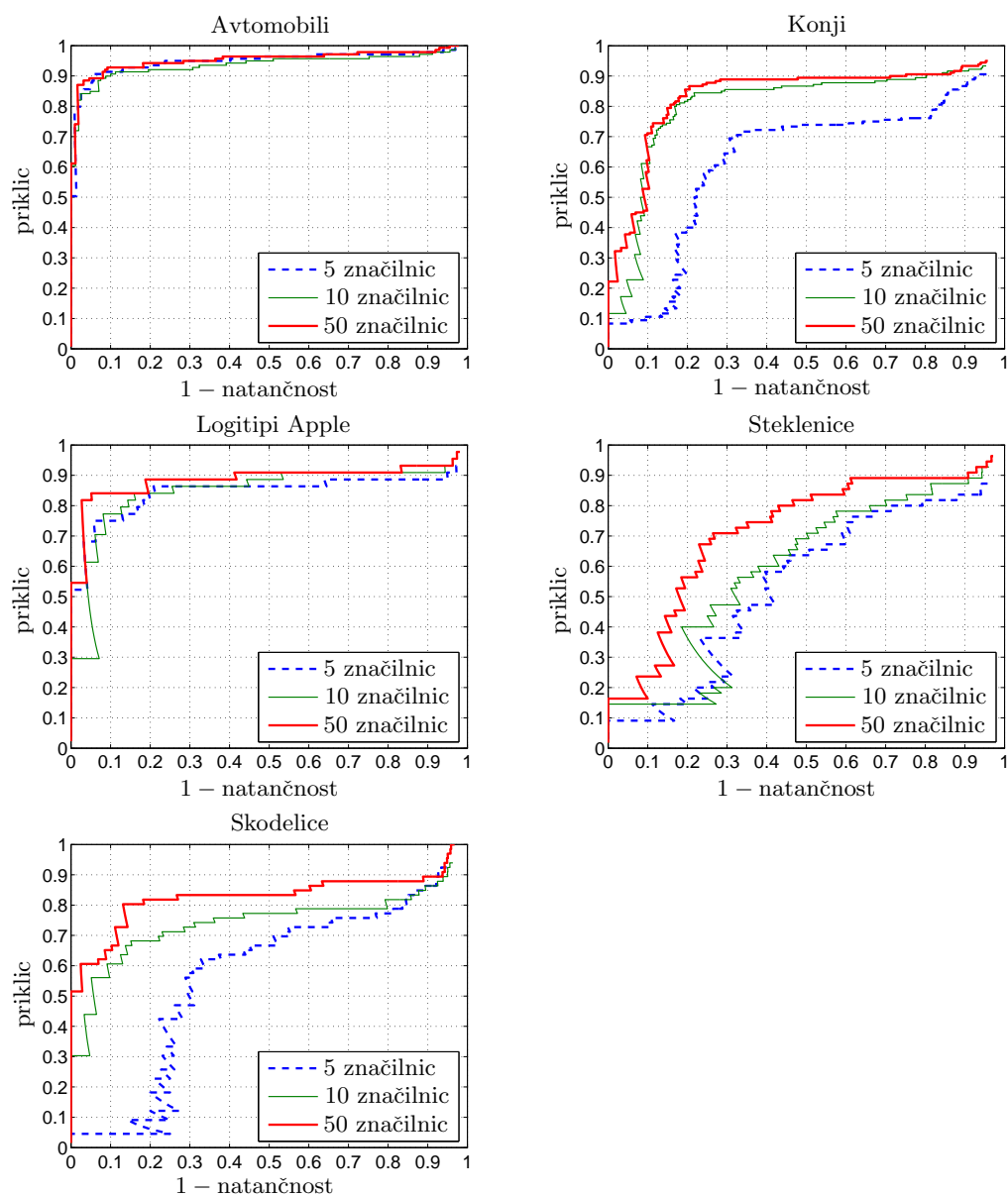
Metoda	F_{\max} [%]	EER [%]
Agarwal in sod. (2004)	44,0	39,6
Fritz in sod. (2005)	Ni podano	87,8
Mutch in Lowe (2006)	Ni podano	90,6

Na sliki 5.8 predstavljamo krivulje RPC za prvih 5, 10 in 50 izbranih značilnic.

Kot je razvidno na sliki 5.7, je možno nadvse vzpodbudne detekcijske rezultate doseči že z zelo majhnim številom izbranih značilnic. Pri avtomobilih smo zgolj s štirimi izbranimi značilnicami dosegli $F_{\max} = 91,0\%$ in $EER = 90,7\%$. Krivulja $F_{\max}(T)$ za $T \in \{1, \dots, 50\}$ zavzame maksimalno vrednost pri sedmih izbranih značilnicah, kjer velja $F_{\max} = 93,7\%$ in $EER = 92,1\%$. Tabela 5.7 podaja detekcijske rezultate, ki jih za zbirko UIUC navajajo drugi raziskovalci. Že pri štirih izbranih značilnicah je dobljena vrednost EER približno enaka rezultatu, ki ga podajajo Mutch in Lowe (2006), sedem izbranih značilnic pa vodi do še višje vrednosti EER.




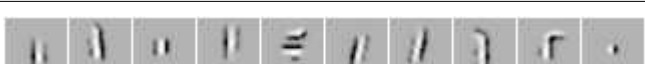

Izkaže se, da je tudi pri drugih slikovnih zbirkah možno z majhnim številom značilnic doseči visoko detekcijsko točnost. Pri logotipih Apple šest izbranih značilnic zadošča za $F_{\max} = EER = 86,4\%$. Pri konjih devet izbranih značilnic vodi do $F_{\max} = 80,5\%$ in $EER = 80,0\%$. Pri skodelicah in steklenicah je za doseg visoke detekcijske točnosti potrebnih nekoliko več značilnic. Pri skodelicah je meja $F_{\max} = 80\%$ prvič presežena pri 14 izbranih značilnicah; pri steklenicah je najboljši rezultat dosežen pri 13 izbranih značilnicah, in sicer $F_{\max} = 75,9\%$ in $EER = 74,6\%$. Karakteristike F_{\max} in EER za posamične vrednosti T so zbrane v ustreznih odsekih tabel 5.17 (F_{\max}) in 5.18 (EER). Kot navajata tabeli, se izbira značilnic izplača pri vseh petih kategorijah: detekcijska točnost, dobljena z uporabo vseh pridobljenih značilnic (njihovo število znaša od 592 do 920), je povsod manjša od detekcijske točnosti pri 50 izbranih značilnicah, pri avtomobilih in logotipih Apple pa že 5 izbranih značilnic vodi do boljših rezultatov kot celotna množica pridobljenih značilnic.

Število izbranih značilnic močno vpliva na porabo časa v testni fazi. Pri detekciji avtomobilov v testni zbirki UIUC potrebuje detekcijski postopek v povprečju 3,9 sekunde za obdelavo ene slike, če njegove odločitve slonijo na prvih petih izbranih značilnicah. Pri uporabi prvih 50 izbranih značilnic obdelava ene slike pri vseh velikostih piramide v povprečju zahteva 16,6 sekunde. Če detekcijski postopek pri tvorbi hipotez uporablja vseh 920 pridobljenih značilnic, povprečni čas obdelave slike naraste na 117,8 sekunde. Časi so bili izmerjeni z izvajanjem naše Matlabove implementacije na računalniku z 2,13-gigaherčnim procesorjem Intel Core 2 Duo. Poraba časa ne narašča linearno z naraščanjem števila značilnic, in sicer zaradi glajenja hipoteznega prostora, pri katerem je poraba časa razmeroma neodvisna od števila izbranih značilnic. Kljub temu moremo trditi, da izbira značilnic občutno



Slika 5.8: Krivulje RPC za prvih 5, 10 oziroma 50 značilnic Fidler in sod., izbranih s filtrsko metodo.

Tabela 5.8: Prvih deset značilnic Fidler in sod., izbranih po postopku AdaBoost.

Kategorija	Prvih 10 izbranih značilnic
Avtomobili	
Konji	
Logotipi Apple	
Steklenice	
Skodelice	

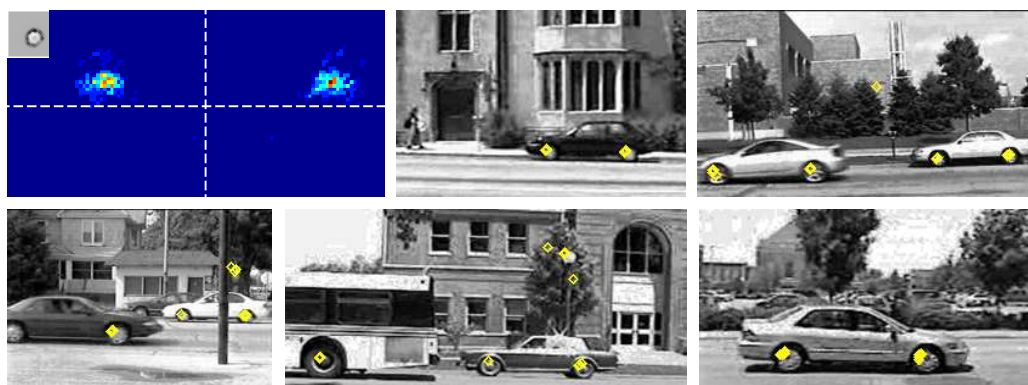
priporočene k razbremenitvi detekcijskega postopka.

Kot prikazuje tabela 5.8, so izbrane značilnice večinoma zelo preproste. Prva izbrana značilnica pri kategoriji avtomobilov se ponovljivo (robustno) odziva na avtomobilska kolesa (slika 5.9a). To je tudi razlog za presenetljivo visoko točnost na precej zahtevni zbirki UIUC. Kolesna značilnica je razmeroma distinktivna in močno prediktivna, saj je na podlagi njene manifestacije mogoče precej zanesljivo sklepati na središče avtomobila. Če detekcija temelji izključno na prvoizbrani kolesni značilnici, detekcijska točnost znaša kar $F_{\max} = 73,1\%$ oz. $EER = 69,1\%$.

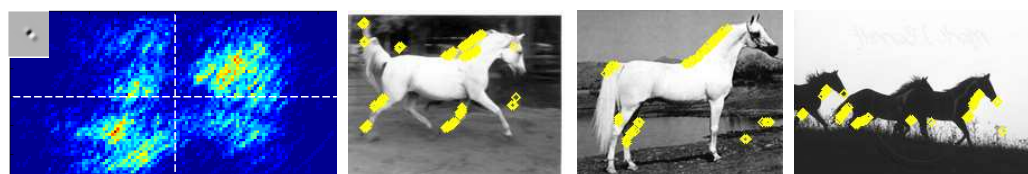
Na sliki 5.9 prikazujemo naučene verjetnostne gostote $p(\mathbf{c} | Z, \mathbf{0})$ za prvoizbrane značilnice za posamezne kategorije, pa tudi primere nastopov prvoizbranih značilnic na testnih slikah. Opazimo, da se značilnice dokaj ponovljivo manifestirajo na vizualno podobnih lokalnih strukturah. Nekatere izbrane značilnice sicer izkazujejo nizko stopnjo specifičnosti, saj se pogosto pojavljajo tudi na ozadju². Vendar pa *množice* izbranih značilnic delujejo sinergično. Prvi dejavnik, ki prispeva k sinergičnosti dane izbirne množice, je postopek AdaBoost, ki skuša odkriti čimbolj medsebojno neodvisen nabor značilnic. Ta lastnost postopka AdaBoost se najnazorneje manifestira pri izbranih značilnicah za skodelice (tabela 5.8): značilnice $Z^{[1]}$, $Z^{[2]}$ in $Z^{[5]}$ pokrivajo različne dele ročaja skodelice. Postopek AdaBoost redundance sicer ne zmore povsem izločiti, saj sta si npr. značilnici $Z^{[2]}$ in $Z^{[8]}$ vizualno zelo podobni in se tudi odzivata na vizualno podobne lokalne strukture.

Drugi dejavnik, ki prispeva k sinergičnosti izbirne množice, je detekcijski postop-

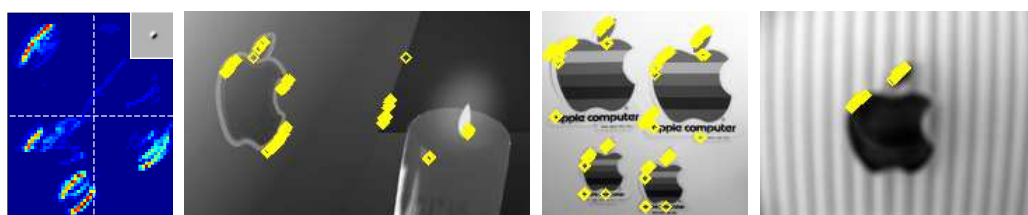
²Za *specifično* značilnico velja, da v testnih slikah nastopa kvečjemu na ugodnih položajih, torej na takšnih, na podlagi katerih je mogoče tvoriti pravilne detekcijske hipoteze. Po tej neformalni definiciji je specifična tudi značilnica, ki se v testnih slikah sploh ne pojavlja. Naj opozorimo, da gre za našo lastno definicijo, saj pojem specifičnosti ni uveljavljen.



(a) Avtomobili



(b) Konji

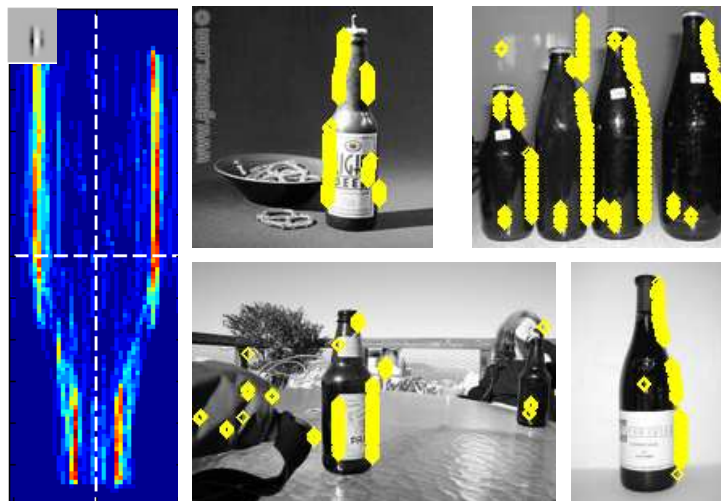


(c) Logotipi Apple (slika se nadaljuje na naslednji strani)

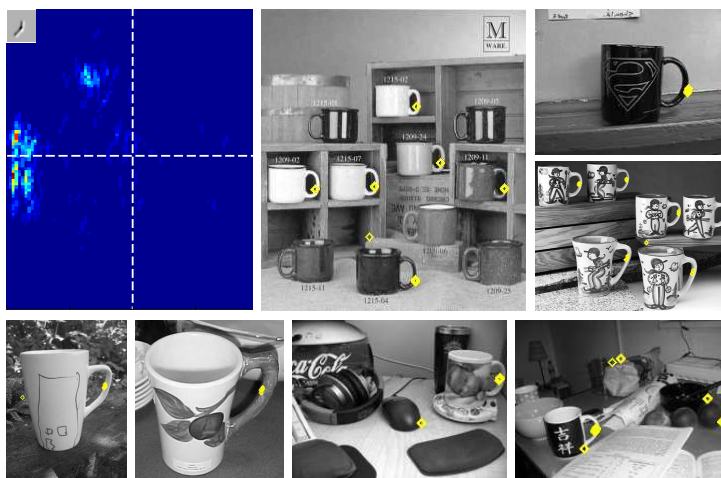
pek, temelječ na implicitnem modelu oblike. *Kombinacija* značilnic lahko omogoča dobre detekcijske rezultate, četudi posamezne značilnice niso pretirano specifične. Če značilnice razmeroma ponovljivo nastopajo na določenih lokalnih strukturah na objektih, se njihovi glasovi za pravilne hipoteze seštevajo (sinergija!), medtem ko glasovi za napačne hipoteze, ki izvirajo iz manifestacij značilnic na ozadju, ponavadi ostajajo razpršeni po hipoteznem prostoru.

Detekcijski mehanizem sicer lahko ublaži negativen vpliv napačno izbranih značilnic, vendar pa ne more vselej nadomestiti neustrezne izbire. Na podlagi povsem nespecifične značilnice ni mogoče tvoriti zanesljivih pravih hipotez, četudi bi se ta značilnica ponovljivo manifestirala na vizualno podobnih lokalnih strukturah. Lahko se tudi zgodi, da vključitev tovrstne značilnice v sicer kako-vostno izbirno množico močno poslabša detekcijske rezultate. Primer nenadnega poslabšanja detekcijske točnosti po izbiri dveh neustreznih značilnic je prikazan na diagramu Avtomobili na sliki 5.15, ki se nanaša na notranjeoptimizacijsko izbiro značilnic Fidler in sod.

Na sliki 5.10 prikazujemo nekaj primerov pravih, napačnih in manjkajočih



(d) Steklenice



(e) Skodelice

Slika 5.9: Prva slika v vsaki skupini prikazuje naučeno aproksimacijo verjetnostne gostote $p(\mathbf{c} | Z^{[1]}, \mathbf{0})$, kjer $Z^{[1]}$ označuje značilnico Fidler in sod., ki jo postopek AdaBoost izbere kot prvo. Ostale slike v posameznih skupinah prikazujejo primere manifestacij teh značilnic na testnih slikah.

detekcijskih hipotez pri optimalnem pragu τ za posamezne testne zbirke. Vse detekcijske rezultate smo pridobili z uporabo prvih 10 značilnic Fidler in sod., izbranih z algoritmom AdaBoost. Naj spomnimo, da je optimalni prag τ določen z enačbo (3.23) in se v skladu z definicijo krivulje RPC izračuna za vsak detekcijski problem oz. testno zbirko posebej.

Slika 5.10 prikazuje poleg številnih pravih detekcij tudi nekaj značilnih lažnih alarmov in manjkajočih detekcij. Manjkajoča detekcija ne pomeni nujno, da detekcijski postopek ni tvoril ustrežajoče pravilne hipoteze. Pogosto je pravilna hipoteza bila zaznana, vendar je njena moč manjša od optimalnega praga τ in se zato takšna hipoteza pri računanju F_{\max} šteje kot manjkajoča detekcija.

Pogost vir lažnih alarmov so grobo teksturirana območja. Manjkajoče detekcije pa so velikokrat posledica slabega sivinskega kontrasta, ki negativno vpliva na količino in ponovljivost nastopov značilnic Fidler in sod. Zanimivo pa je, da je v nekaterih redkih primerih za nastop manjkajoče detekcije odgovoren postopek MDL. Osredotočimo se na sliko, ki je prikazana na levi strani v drugi vrsti na sliki 5.10a. Slika prikazuje lažni alarm in hkrati tudi manjkajočo detekcijo, saj temni avtomobil pri optimalnem pragu τ ni bil detektiran. Slika 5.11 prikazuje hipotezni prostor po zaključenem glasovanju in glajenju z Gaussovim filtrom z velikostjo 11×11 in z vrednostjo parametra $\sigma = 2$ (tak filter smo uporabili pri vseh kategorijah). Hipoteza H_1 je najmočnejša in ustreza označenemu lažnemu alarmu. Po moči ji sledi hipoteza H_2 , ki ustreza pravilni detekcijski hipotezi. Sprednje kolo temnega avtomobila botruje nastanku šibkejše hipoteze H_3 , sprednje kolo svetlega avtomobila pa je vzrok za nastanek še šibkejše hipoteze H_4 . Brez uporabe postopka MDL bi najbrž obe najmočnejši hipotezi (H_1 in H_2) presegli optimalni prag. Postopek MDL pa v prvi iteraciji izbere napačno hipotezo H_1 in nato zaradi skupnih komponent vzvratnih projekcij hipotez H_1 in H_2 — gre za zadnje kolo temnega avtomobila, ki zelo veliko prispeva k obema hipotezama — v precejšnji meri zaduši pravilno hipotezo H_2 (in seveda tudi H_4 , kar pa je za razliko od dušenja H_2 dobrodošlo). Tako moč H_2 pade globoko pod optimalni prag τ . Postopek MDL je tudi povzročitelj izpada detekcije ene od steklenic na drugi sliki na sliki 5.10d.

Kljub temu se postopek MDL izkaže kot izjemno pomembna pridobitev. Tabela 5.9 primerja detekcijske rezultate, dobljene z uporabo postopka MDL, in rezultate, dobljene brez uporabe tega postopka. Tabela prikazuje rezultate pri 50 izbranih značilnicah in za filtrsko izbirno metodo, izkaže pa se, da postopek MDL v vseh primerih znatno izboljša detekcijsko točnost. Upoštevati je namreč treba, da postopek MDL duši hipoteze vzdolž celotne velikostne piramide in da posledično oslabi marsikatero hipotezo, ki bi bila sicer proglašena za lažni alarm zaradi večkratne (četudi pravilne) detekcije istega objekta.

5.3.2 Delovanje postopka AdaBoost

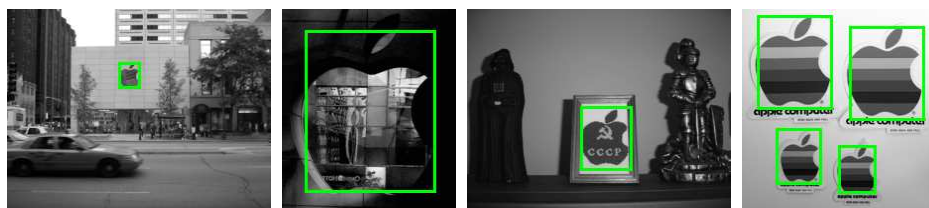
V nadaljevanju se bomo še nekoliko pomudili pri izbirnem postopku AdaBoost. Uporabljali bomo notacijo, ki smo se je poslužili v algoritmu 2.1. Kot je razvi-



(a) Avtomobili



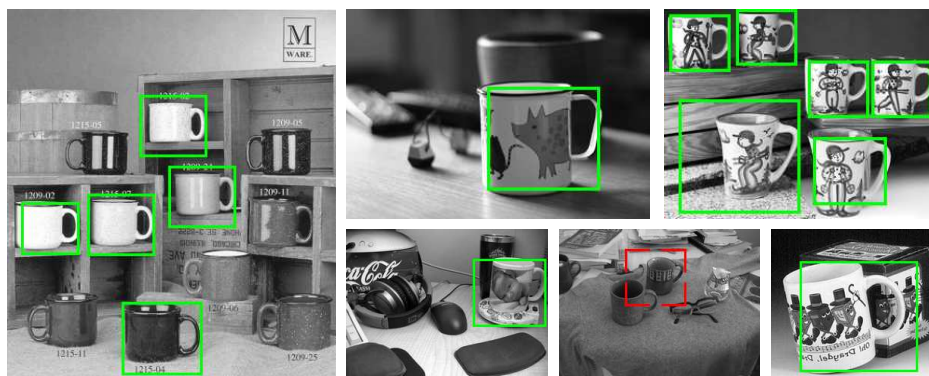
(b) Konji



(c) Logotipi Apple

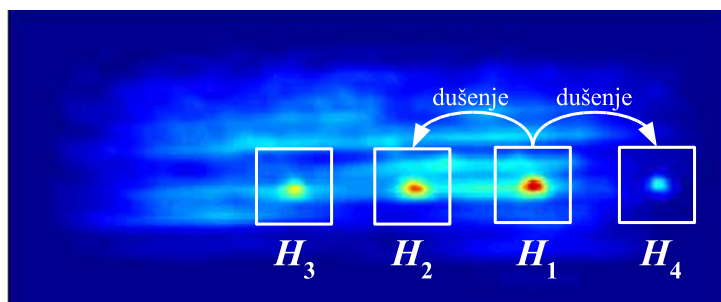


(d) Steklenice



(e) Skodelice

Slika 5.10: Primeri pravih, napačnih in manjkajočih detekcij pri optimalnem pragu τ krivulje RPC. Detekcijski postopek je pri vsaki posamezni kategoriji uporabljal prvih 10 značilnic Fidler in sod., izbranih s postopkom AdaBoost. Prva slika v sklopu primerov za zbirko skodelic vsebuje le eno manjkajočo detekcijo, saj skodelice z ročajem na levi strani niso obravnavane kot objekti.

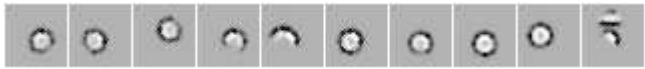






Slika 5.11: Zglajeni glasovalni prostor za prvo sliko v drugi vrsti slike 5.10a. Pravilna hipoteza H_2 je nekoliko šibkejša od lažnega alarma H_1 .

Tabela 5.9: Primerjava detekcijskih rezultatov, pridobljenih z uporabo postopka MDL, in rezultatov, pridobljenih brez uporabe tega postopka. Rezultati so podani kot $100 \times F_{\max}$ in se nanašajo na izbor prvih 50 značilnic po filterski izbirni metodi. Preizkusili smo oba sestava značilnic, Fidler in sod. ter Ferrari in sod.

Kategorija	Fidler in sod.		Ferrari in sod.	
	MDL	Brez MDL	MDL	Brez MDL
Avtomobili	92,5	63,4	61,7	46,2
Konji	83,0	35,9	71,7	34,6
Logotipi Apple	89,2	76,1	86,1	77,6
Steklenice	72,2	31,1	66,7	60,2
Skodelice	83,5	43,0	76,2	59,5

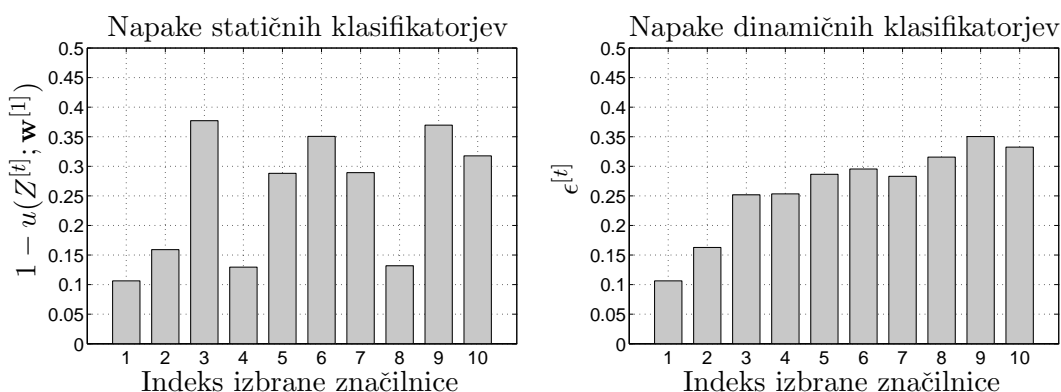
Tabela 5.10: Prvih deset značilnic Fidler in sod., izbranih po padajočih vrednostih $u(Z; \mathbf{w}^{[1]})$.

Kategorija	Prvih 10 izbranih značilnic
Avtomobili	
Konji	
Logotipi Apple	
Steklenice	
Skodelice	

dno na sliki 5.7, izbira značilnic po postopku AdaBoost omogoča bistveno boljše detekcijske rezultate kot izbira po padajočih statičnih ocenah $u(Z; \mathbf{w}^{[1]})$. Tabela 5.10 prikazuje prvih deset značilnic, izbranih po padajočih vrednostih $u(Z; \mathbf{w}^{[1]})$. Očitno je, da tako izbrane množice značilnic vsebujejo veliko stopnjo redundance, saj so vizualno podobne značilnice tudi podobno ocenjene. Na primer, prvih deset značilnic, izbranih glede na $u(Z; \mathbf{w}^{[1]})$ za kategorijo skodelic, pokriva le spodnji del ročaja; značilnice, ki pokrivajo zgornji del ročaja, so prejele nižje statične ocene.

Levi diagram na sliki 5.12 prikazuje stopnje napake *statičnih* optimalnih klasifikatorjev za prvih 10 značilnic, izbranih s postopkom AdaBoost za kategorijo skodelic. Prag statičnega optimalnega klasifikatorja je nastavljen glede na izhodiščne vrednosti slikovnih uteži, $\mathbf{w}^{[1]}$, klasifikatorjeva stopnja napake pa znaša $1 - u(Z^{[t]}; \mathbf{w}^{[1]})$. Desni diagram slike 5.12 prikazuje stopnje napake *dinamičnih* optimalnih klasifikatorjev istih značilnic. Prag dinamičnega optimalnega klasifikatorja se nastavi glede na uteži slik \mathcal{I}_i v trenutni iteraciji postopka AdaBoost. Stopnja napake dinamičnega klasifikatorja znaša $\epsilon^{[t]} = 1 - u(Z^{[t]}; \mathbf{w}^{[t]})$, kjer t podaja indeks AdaBoostove iteracije, $Z^{[t]}$ značilnico, izbrano v iteraciji t , $\mathbf{w}^{[t]}$ pa vektor slikovnih uteži v iteraciji t .

Ker se v prvi iteraciji postopka AdaBoost uporabljajo izhodiščne uteži slik \mathcal{I}_i , je prvoizbrana značilnica najboljša tako v smislu statične ocene ($u(Z^{[1]}; \mathbf{w}^{[1]})$) kot v smislu dinamične ocene ($1 - \epsilon^{[1]}$). Stopnja napake ustreznega statičnega klasifikatorja je zato enaka stopnji napake ustreznega dinamičnega klasifikatorja. Vendar pa se statična in dinamična stopnja napake že v tretji iteraciji postopka AdaBoost močno razlikujeta. Dinamični klasifikator je skupaj s pripadajočo značilnico izbran tako, da doseže najnižjo stopnjo napake pri trenutnih vrednostih uteži. Ker



Slika 5.12: Diagram na levi ponazarja stopnje napake statičnih optimalnih klasifikatorjev prvih 10 značilnic Fidler in sod., izbranih s postopkom AdaBoost za kategorijo skodelic. Diagram na desni predstavlja stopnje napake dinamičnih optimalnih klasifikatorjev istih izbranih značilnic.

Tabela 5.11: Tabela za vsako od značilnic, izbranih v prvih 10 iteracijah postopka AdaBoost, podaja njen indeks v padajočem zaporedju $u(Z; \mathbf{w}^{[1]})$. Podatki se nanašajo na značilnice Fidler in sod. ter na kategorijo skodelic.

Indeks izbrane značilnice	1	2	3	4	5	6	7	8	9	10
Indeks v pad. zaporedju $u(Z; \mathbf{w}^{[1]})$	1	7	156	3	28	109	30	4	142	49

pa se uteži slik v tretji iteraciji že bistveno razlikujejo od izhodiščnih uteži, je razlika med statično in dinamično oceno značilnice precej velika.

Dinamična stopnja napake narašča skoraj monotonno. Če postopek AdaBoost, prikazan kot algoritem 2.1, ne bi sproti odstranjeval izbranih značilnic, ampak bi lahko tudi večkrat izbral isto značilnico, bi stopnja napake naraščala povsem monotonno (Freund in Schapire, 1997).

Tabela 5.11 na nekoliko drugačen način ponazarja razliko med izbiro v okviru postopka AdaBoost in izbiro po padajočih statičnih ocenah. Značilnica, izbrana v tretji AdaBoostovi iteraciji za kategorijo skodelic, na primer, v padajočem vrstnem redu statičnih ocen zaseda šele 156. mesto (od 912), kar priča o velikem neskladju med obema načinoma izbiranja značilnic.

Diagrami na slikah 5.13 in 5.14 natančneje osvetljujejo dogajanje v prvih treh iteracijah postopka AdaBoost pri izbiri značilnic za kategorijo skodelic. Diagrami na sliki 5.13 ponazarjajo točnosti prvih treh izbranih dinamičnih optimalnih klasifikatorjev na vseh 408 (= 184 pozitivnih + 224 negativnih) učnih slikah \mathcal{I}_i . Prisotnost navpične črte pri indeksu slike pomeni, da dinamični optimalni klasifikator pravilno uvrsti dano sliko (glede na optimalni prag $\theta^{[t]}$), odsotnost črte pa ponazarja napačno klasifikacijo dane slike.

Drugoizbrani dinamični klasifikator skoraj idealno dopolnjuje prvoizbranega. Prvoizbrani klasifikator namreč napačno uvrsti 42 slik, drugoizbrani klasifikator

pa od teh pravilno uvrsti kar 36 slik. Število slik, ki jih napačno uvrstita oba klasifikatorja, je torej enako 6. Tretjeizbrani klasifikator, kot je razvidno na ustreznem diagramu, se sicer odreže znatno slabše na negativnih učnih slikah, vendar pravilno uvrsti kar pet od šestih slik, pri katerih sta se zmotila prva dva izbrana klasifikatorja. Tretjeizbrani klasifikator torej odlično dopolnjuje prvi dve izbiri postopka AdaBoost. Komplementarnost izbranih klasifikatorjev se odseva tudi v komplementarnosti izbranih značilnic: tretjeizbrana značilnica pokrije določen delež „informacije“, ki je ni pokrila nobena od prvih dveh izbranih značilnic.

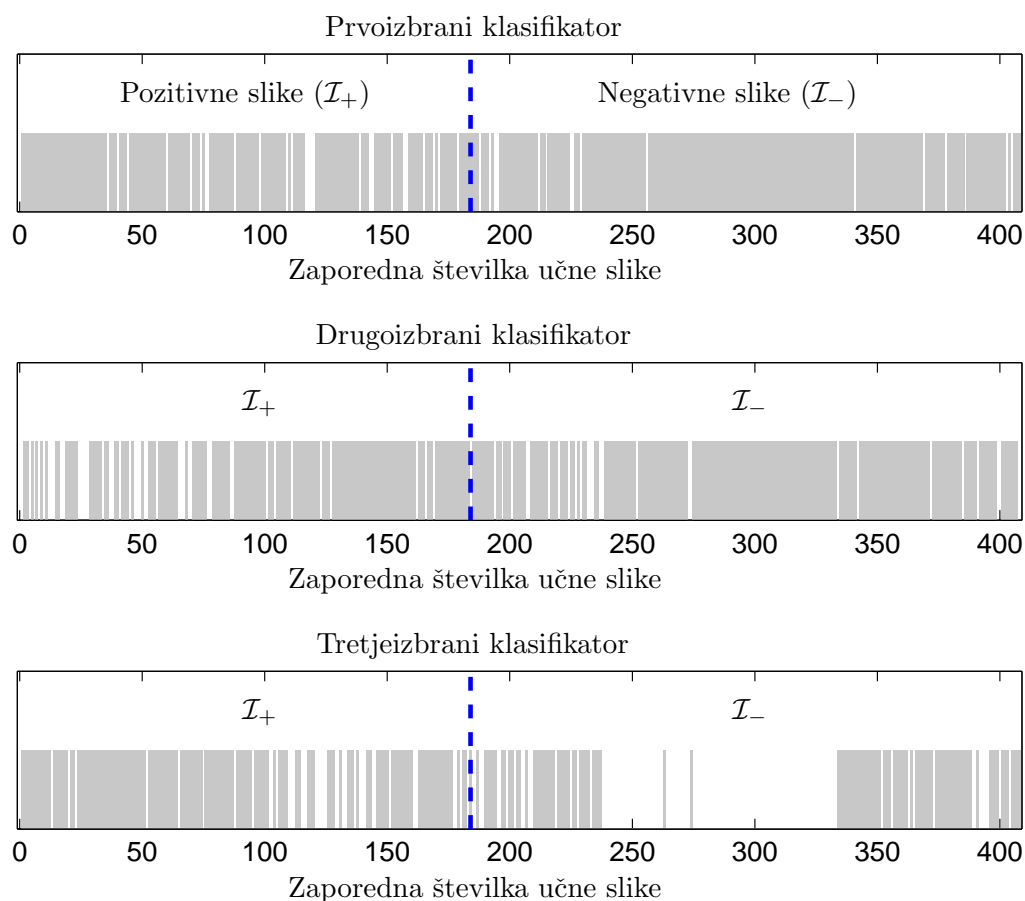
Diagrami na sliki 5.14 razkrivajo spreminjanje slikovnih uteži v prvih štirih AdaBoostovih iteracijah. Po koncu vsake iteracije se povečajo relativne uteži slik, ki jih je nazadnje izbrani klasifikator napačno klasificiral. Na tretjem diagramu izstopa šest vrhov, ki ustrezajo tistim šestim slikam, ki sta jih napačno uvrstila oba klasifikatorja, izbrana v prvih dveh iteracijah. Na četrtem diagramu izstopa utež slike z indeksom 111, ki je ni pravilno klasificiral nobeden od prvih treh dinamičnih optimalnih klasifikatorjev.

5.4 Rezultati notranjeoptimizacijske izbire značilnic Fidler in sod.

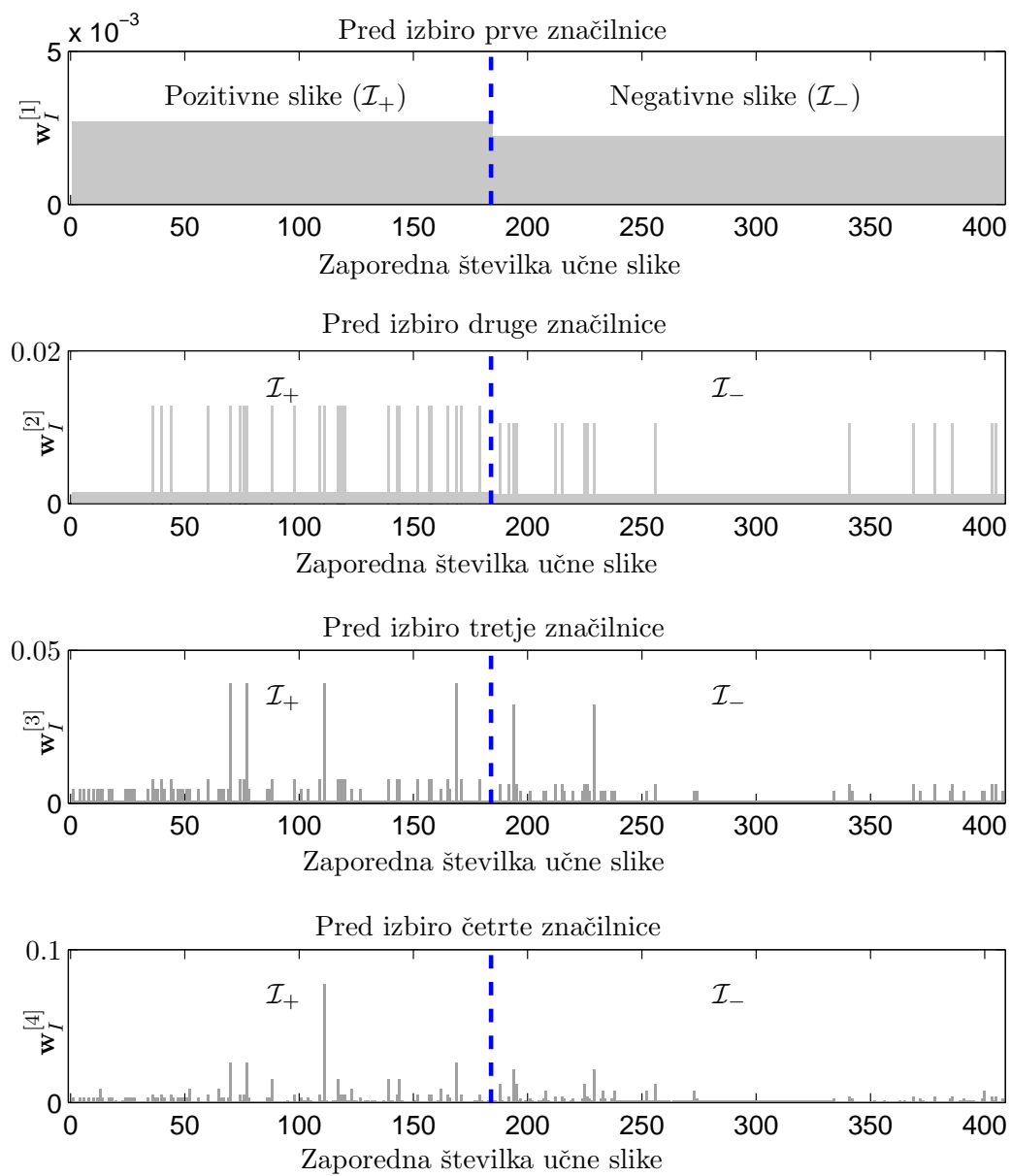
Omenili smo, da je poglobljena slabost notranjeoptimizacijskega pristopa k izbiri značilnic velika časovna zahtevnost. Zaradi tega smo bili primorani uporabiti razmeroma majhno množico učnih slik \mathcal{I}_i in privzeti določene kompromisne predpostavke in poenostavitve.

Učne slike \mathcal{I}_i sicer prikazujejo raznolike objekte na precej različnih ozadjih, vendar pa je celokupna raznolikost množice \mathcal{I}_i zaradi zelo omejenega števila učnih slik premajhna, da bi notranjeoptimizacijski pristop mogel ustvariti enako kakovosten izbor značilnic kot filterski pristop. Kot je razvidno na sliki 5.15, ki prikazuje odvisnost F_{\max} od števila izbranih značilnic, je notranjeoptimizacijski pristop „konkurenčen“ filterskemu le pri razmeroma enostavni kategoriji logotipov Apple. Pri drugih kategorijah so rezultati notranjeoptimizacijske izbire bistveno slabši.

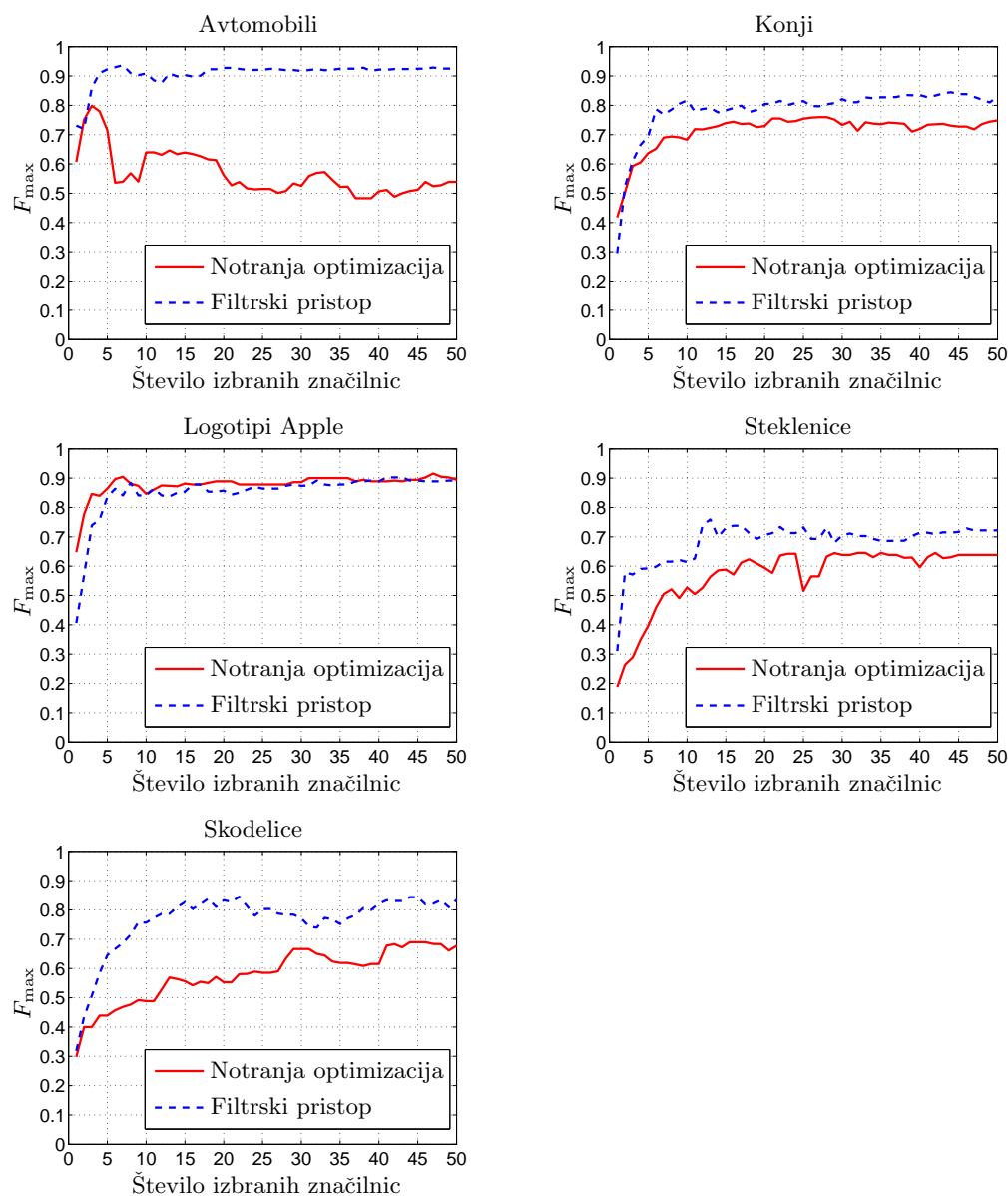
Medtem ko krivulje $F_{\max}(T)$ za filtersko izbiro ne izkazujejo izrazitih padcev, je pri notranjeoptimizacijski metodi žal drugače. Posebej očiten je strm zdrs detekcijske točnosti po obetavnem začetku na testni množici kategorije avtomobilov. Zaradi premajhnega števila slik \mathcal{I}_i se namreč praviloma zgodi, da v dani iteraciji požrešnega izbirnega postopka več različnih kandidatnih značilnic doseže isto kriterijsko vrednost $(F_{\max} + \text{EER})/2$ na množici \mathcal{I}_i . V takem primeru se značilnica izbere naključno. Značilnici, ki se pri avtomobilih izbereta v peti in šesti iteraciji, sicer ustrezata požrešnemu kriteriju na učni množici, vendar pa isto vrednost $(F_{\max} + \text{EER})/2$ doseže tudi več drugih značilnic. Kljub izkazani koristnosti za detekcijo objektov na *učni* množici se značilnici, izbrani v 5. in 6. iteraciji, na *testni* množici izkažeta kot zelo slabi izbiri, saj sta povsem nedistinktivni in neprediktivni. Krivulja $F_{\max}(T)$ se po izbiri teh značilnic ne more več bistveno izboljšati.



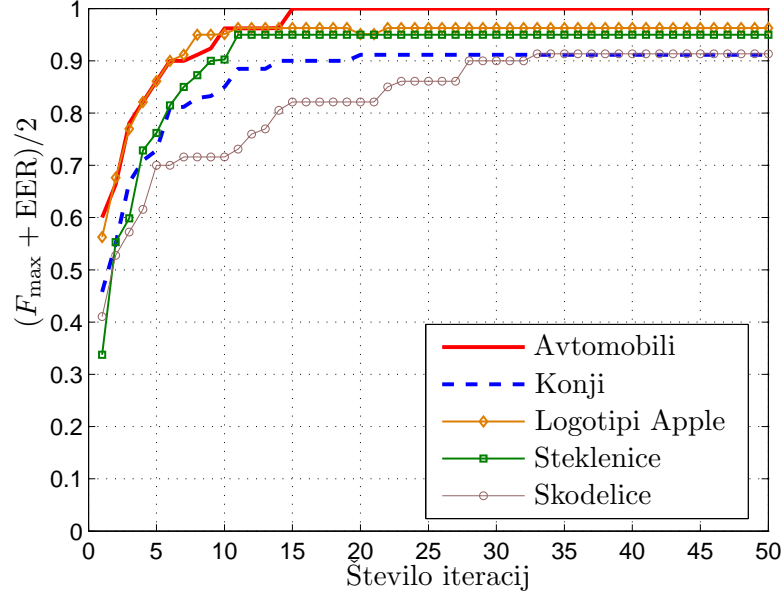
Slika 5.13: Klasifikacijska točnost dinamičnih optimalnih klasifikatorjev, izbranih v prvih treh iteracijah postopka AdaBoost. Prisotnost črte pri slikovnem indeksu predstavlja pravilno klasifikacijo ustrezne slike glede na dinamični optimalni prag, prazen prostor (odsotnost črte) pa napačno klasifikacijo. Podatki se nanašajo na značilnice Fidler in sod. ter na kategorijo skodelic.



Slika 5.14: Spreminjanje uteži učnih slik v prvih štirih iteracijah postopka Ada-Boost. Podatki se nanašajo na značilnice Fidler in sod. ter na kategorijo skodelic.



Slika 5.15: Odvisnost F_{\max} od števila izbranih značilnic Fidler in sod. Diagrami podajajo primerjavo med notranjeoptimizacijsko izbiro (polne rdeče krivulje) in izbiro po metodi AdaBoost (črtkane modre krivulje).



Slika 5.16: Spreminjanje kriterija $(F_{\max} + EER)/2$ na množici \mathcal{I}_i pri požrešnem izbiranju značilnic Fidler in sod. v sklopu notranjeoptimizacijske izbirne metode.

Slika 5.16 prikazuje spreminjanje vrednosti kriterija $(F_{\max} + EER)/2$ na učni množici pri požrešnem izbiranju značilnic. Kot je razvidno na diagramu, kriterij že po 10 do 15 izbranih značilnic doseže „strop“. Edina izjema je kategorija skodelic, kjer je največja vrednost kriterija v okviru prvih 50 iteracij dosežena šele v 33. iteraciji požrešnega postopka.

Tabela 5.12 prikazuje prvih deset izbranih značilnic za vsako posamezno kategorijo. Ker notranja optimizacija upošteva medsebojno odvisnost značilnic, je množica izbranih značilnic bolj raznolika kot množica značilnic, izbranih po padajočih ocenah $u(Z; \mathbf{w}^{[1]})$ (tabela 5.10). Raznolikost izbirne množice pa seveda še ni zadosten pogoj za dejansko uporabnost izbirne množice v fazi detekcije.






Vrednosti mer F_{\max} in EER v odvisnosti od števila izbranih značilnic so zbrane v ustreznih odsekih tabel 5.17 (F_{\max}) in 5.18 (EER).

5.5 Rezultati filtrske izbire značilnic Ferrari in sod.

Kot je razvidno na sliki 5.17 ter v tabelah 5.17 in 5.18, se značilnice Ferrari in sod. pri detekciji objektov v danih testnih zbirkah odrežejo slabše od značilnic Fidler in sod. Za tak rezultat je odgovornih več dejavnikov:

- Reprezentacija slike z množico T izbranih značilnic Ferrari in sod. tipično vsebuje bistveno manj manifestacij kot reprezentacija z množico T izbranih

Tabela 5.12: Prvih deset značilnic Fidler in sod., izbranih po metodi na osnovi notranje optimizacije.

Kategorija	Prvih 10 izbranih značilnic
Avtomobili	
Konji	
Logotipi Apple	
Steklenice	
Skodelice	

značilnic Fidler in sod, kar je posledica načina učenja značilnic in reprezentacijskega postopka. Če se značilnice pojavljajo redkeje, je tudi verjetnost njihovega ponovljivega nastopanja manjša.

Kot prikazuje slika 5.9, značilnice Fidler in sod. izkazujejo visoko stopnjo ponovljivosti (npr. prvoizbrana značilnica za skodelice se ponovljivo manifestira na spodnjem delu ročajev skodelic), čeprav povečini niso pretirano specifične (omenjena značilnica se ne pojavlja le na ročajih, temveč tudi na drugih podobnih lokalnih strukturah). Vendar je, kot smo omenili, detekcijska metoda zasnovana tako, da nagrajuje ponovljive (robustne) značilnice, četudi so te razmeroma nespecifične. Za zmanjšanje učinka nizke stopnje specifičnosti deloma poskrbi že osnovni detekcijski mehanizem, ki omogoči koncentracijo glasov za pravilne hipoteze pod pogojem ponovljivega nastopanja izbranih značilnic, deloma pa tudi algoritem MDL, ki dodatno zaduši nekatere napačne hipoteze. Če lokalnih struktur ni mogoče ponovljivo zaznavati, pa ostaneta tako detekcijski mehanizem kakor tudi algoritem MDL brez moči, saj detekcijske hipoteze izvirajo iz manifestacij značilnic in jih ni mogoče ustvariti iz ničesar.

Pogostost nastopanja značilnic Ferrari in sod. v slikah bi sicer bilo mogoče povečati z dvigom praga t , ki uravnava reprezentiranje skupin k robnih odsekov z značilnicami. Vendar bi dvig praga povečal tudi možnost, da se neki skupini priredi ena ali več vizualno neustreznih značilnic, kar bi izničilo pozitiven učinek višje stopnje ponovljivosti. Izbrane značilnice Fidler in sod. so povečini res precej nespecifične, vendar pa se vedno odzivajo na vizualno

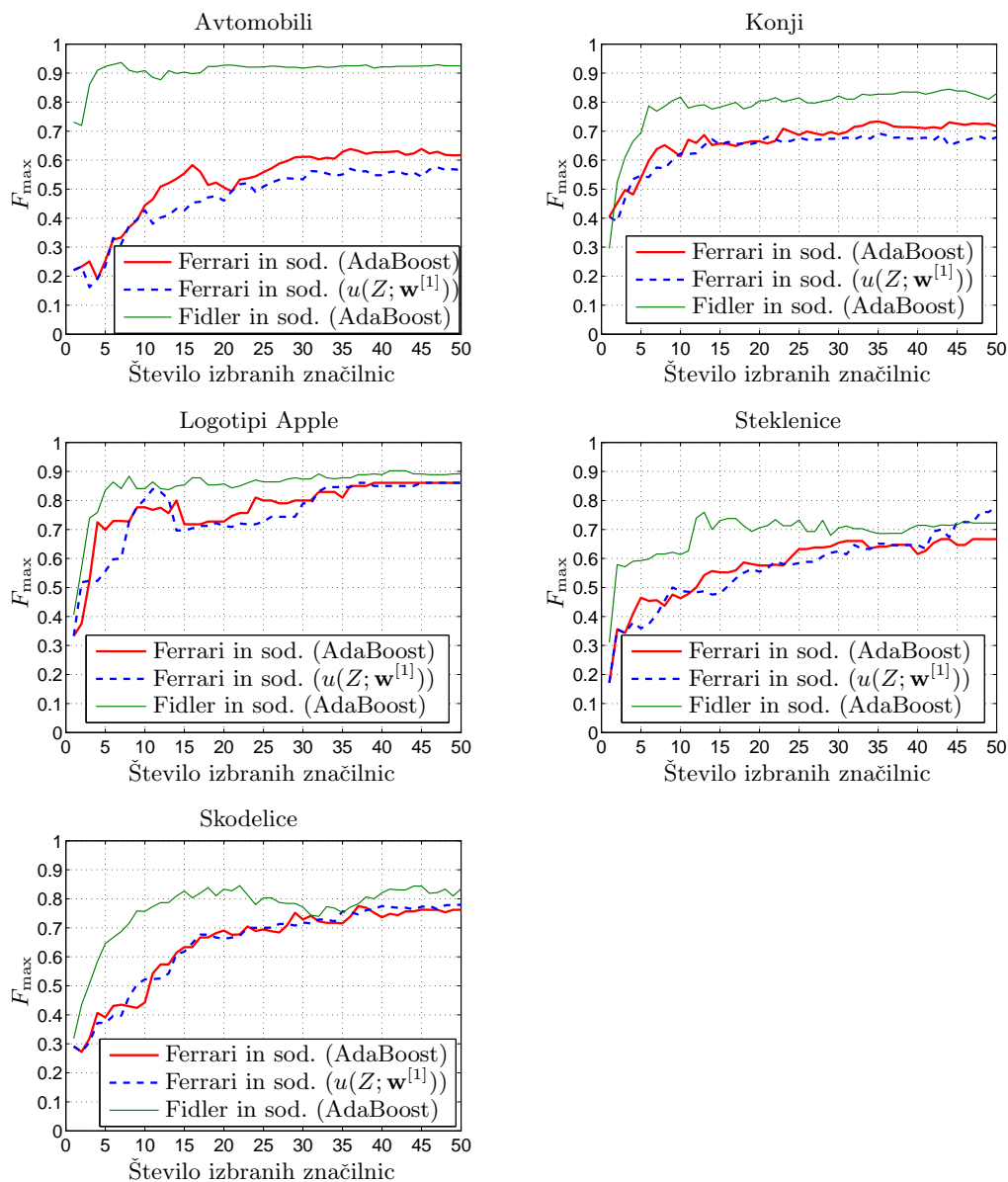
podobne strukture.

- Značilnice Ferrari in sod. ne morejo dobro zajeti krožnih in ločnih lokalnih struktur, ki so pri detekciji nekaterih kategorij (avtomobili, skodelice) zelo pomembne. Krog je mogoče zajeti zgolj z velikim številom ravnih odsekov, pa še to le pri dovolj velikem premeru. Povezan par ravnih odsekov je zmožen — v zelo omejenem obsegu — zajeti kvečjemu krožni lok.
- Postopek reprezentiranja slike z značilnicami Fidler in sod. je določen s samo zgradbo značilnic. Ker so značilnice definirane kot fleksibilne kompozicije, je reprezentacija precej robustna v smislu občutljivosti na manjše deformacije lokalnih struktur. To je tudi razlog za ponovljivo nastopanje značilnic Fidler in sod. na slikah. Postopek reprezentiranja slike z značilnicami ravni n je praktično povsem neparametričen, saj je skoraj vsa informacija zajeta v sami strukturi značilnic in v že obstoječi reprezentaciji slike z ravno $n - 1$.

Pri značilnicah Ferrari in sod. so razmere nekoliko drugačne. Reprezentacija slike z značilnicami temelji na *ad hoc* meri razdalje (enačba 4.6) in na podobnostnem pragu. Mera razdalje je odvisna od treh uteženih dejavnikov. Skupno štirje reprezentacijski parametri (tri uteži in prag) močno vplivajo na reprezentacijski postopek in s tem tudi na kakovost detekcije. Žal pa določanje najprimernejših vrednosti parametrov še zdaleč ni enostavno. V implementaciji smo uteži skušali nastaviti z opazovanjem gručenja skupin robnih odsekov v značilnice, podobnostni prag pa z izvajanjem detekcijskega postopka na neki učni množici pri različnih kandidatnih vrednostih. Seveda pa tovrstno preizkušanje še zdaleč ni zanesljivo jamstvo za pridobitev optimalnega nabora parametrov.



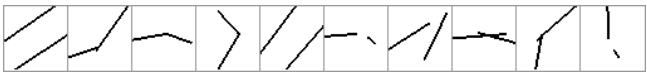

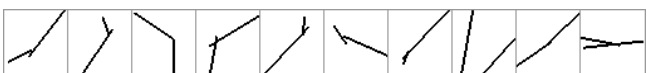
Razlika med detekcijsko sposobnostjo obeh sestavov značilnic je še posebej izstopajoča pri avtomobilih. Pri uporabi sestava Fidler in sod. so za detekcijski uspeh najbolj zaslužne značilnice, ki se ponovljivo odzivajo na avtomobilska kolesa. Sestav Ferrari in sod. takšnih značilnic seveda ne vsebuje, zato se mora detekcijski postopek opreti na druge lokalne strukture podob avtomobilov. Vendar pa je kolo verjetno daleč najbolj distinktivna in prediktivna lokalna struktura za kategorijo avtomobilov, zato izpada robustne zaznave koles ni mogoče enostavno nadomestiti.

Tabela 5.13 prikazuje prvih deset značilnic Ferrari in sod., ki jih za posamezne kategorije izbere postopek AdaBoost. Tabela 5.14 prikazuje prvih deset značilnic, izbranih po padajočem vrstnem redu ocen $u(Z; \mathbf{w}^{[1]})$. Opazimo, da so razlike med odgovarjajočimi izbori razmeroma majhne. Tudi v smislu dobljene detekcijske točnosti (slika 5.17) so razlike med obema načinoma filtrskega izbiranja pri sestavu Ferrari in sod. bistveno manjše kot pri sestavu Fidler in sod. Ker se, kot smo že omenili, značilnice Ferrari in sod. pojavljajo precej redkeje kot značilnice Fidler in sod., za večino kombinacij pozitivnih učnih slik $I \in \mathcal{I}_+ \subset \mathcal{I}_i$ in izhodiščnih značilnic Z velja $S(Z, I) = 0$, kakor narekuje (2.7). Zato je celo stopnja napake klasifikatorja, ki ga AdaBoost izbere v prvi iteraciji, razmeroma velika. Pri skodelicah, na primer, velja $1 - u(Z^{[1]}; \mathbf{w}^{[1]}) = 0,345$. Ker je poleg tega število



Slika 5.17: Odvisnost F_{\max} od števila izbranih značilnic. Na sliki primerjamo izbiro značilnic Ferrari in sod. po metodi AdaBoost (polne rdeče krivulje), izbiro značilnic Ferrari in sod. po padajočih ocenah $u(Z; \mathbf{w}^{[1]})$ (črtkane modre krivulje) in izbiro značilnic Fidler in sod. po metodi AdaBoost (polne zelene krivulje).

Tabela 5.13: Prvih 10 značilnic Ferrari in sod., izbranih s postopkom AdaBoost.

Kategorija	Prvih 10 izbranih značilnic
Avtomobili	
Konji	
Logotipi Apple	
Steklenice	
Skodelice	

izhodiščnih značilnic Ferrari in sod. manjše od števila izhodiščnih značilnic Fidler in sod. (tabela 5.17, stolpca $|\mathcal{Z}|$), ima AdaBoost pri izbiranju značilnic Ferrari in sod. precej manj „manevrskega prostora“ kot pri izbiranju značilnic Fidler in sod.

Dejstva, da se značilnice Ferrari in sod. po detekcijskih rezultatih obnesejo slabše od značilnic Fidler in sod., ne gre tolmačiti kot kritiko značilnic Ferrari in sod. *per se*. Pokazali smo le, da sestav Ferrari in sod. dosega slabše rezultate od sestava Fidler in sod. v okviru detekcijskega postopka na osnovi ISM, dveh različnih izbirnih metod in v okviru preizkusov na omejeni množici vizualnih kategorij. Zato ne moremo trditi, da je en sestav „boljši“ od drugega. To smo poudarili že v razdelku 4.1, ko smo navedli, da so bile značilnice Ferrari in sod. razvite na podlagi drugačnih temeljev in z drugačnimi nameni kot značilnice Fidler in sod.

V posebnem poskusu smo preverili smiselnost odločitve, da se vsaka testna slika obravnava pri treh različnih relativnih velikostih ($1/\sqrt{2}$, 1 in $\sqrt{2}$) in ne zgolj na podlagi izhodiščne velikosti (relativna velikost 1), kar bi pri uporabi velikostno invariantnih značilnic Ferrari in sod. načeloma zadoščalo. Rezultati, zbrani v tabeli 5.15, to odločitev podpirajo, saj je večvelikostna obravnava testnih slik vodila do znatno večje detekcijske točnosti kot enovelikostna obravnava.

Slika 5.18 prikazuje primere nastopov značilnic Ferrari in sod. na testnih slikah.

5.6 Rezultati notranjeoptimizacijske izbire značilnic Ferrari in sod.

Po pričakovanjih so detekcijski rezultati, dobljeni na podlagi notranjeoptimizacijske izbire značilnic Ferrari in sod., še nekoliko slabši od rezultatov, dobljenih na

Tabela 5.14: Prvih 10 značilnic Ferrari in sod., izbranih po padajočih ocenah $u(Z; \mathbf{w}^{[1]})$.

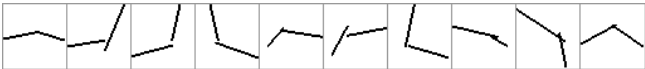
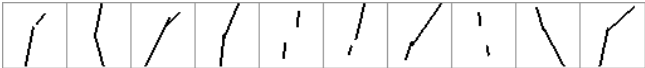
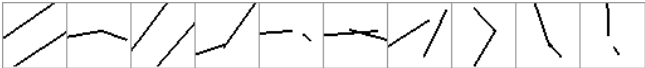


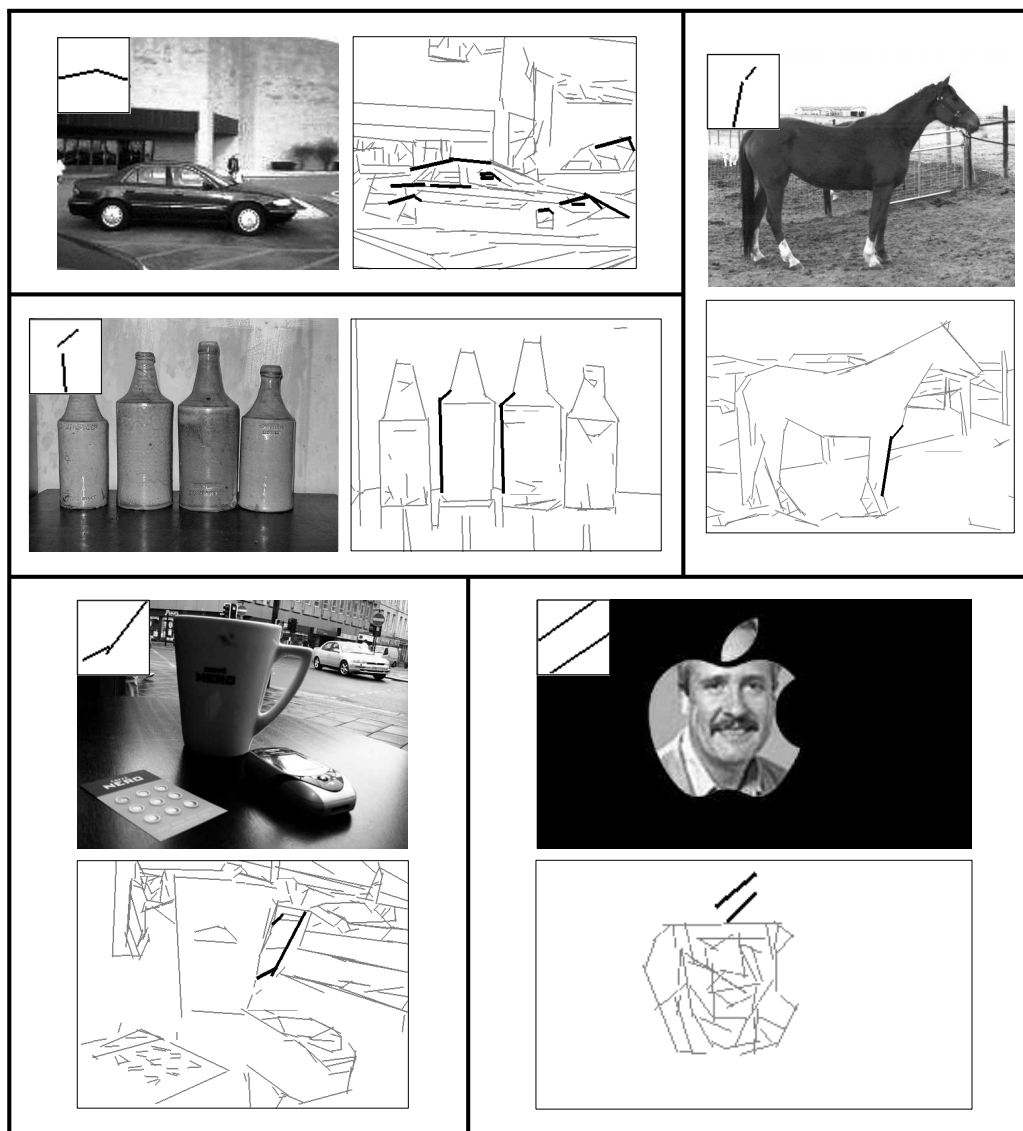
Kategorija	Prvih 10 izbranih značilnic
Avtomobili	
Konji	
Logotipi Apple	
Steklenice	
Skodelice	

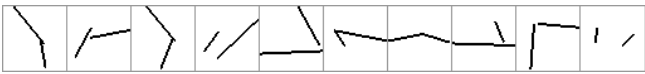
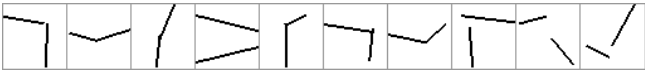
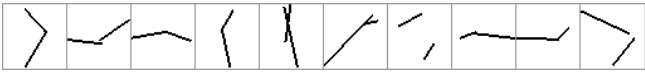
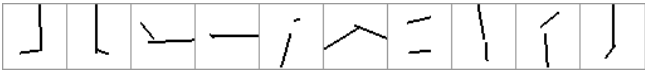
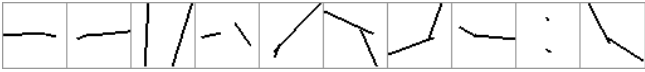
Tabela 5.15: Primerjava detekcijske točnosti, dobljene z obravnavo testnih slik v relativnih velikostih ($1/\sqrt{2}, 1, \sqrt{2}$), in točnosti, dobljene z obravnavo testnih slik v izhodiščni velikosti (relativna velikost 1). Podatki se nanašajo na izbiro značilnic Ferrari in sod. s postopkom AdaBoost. Detekcijske točnosti v tabeli so izražene kot $100 \times F_{max}$.

Kategorija	5 izbranih značilnic		10 izbranih značilnic		50 izbranih značilnic	
	$\{1/\sqrt{2}, 1, \sqrt{2}\}$	$\{1\}$	$\{1/\sqrt{2}, 1, \sqrt{2}\}$	$\{1\}$	$\{1/\sqrt{2}, 1, \sqrt{2}\}$	$\{1\}$
Avtomobili	25,3	18,3	44,4	28,8	61,7	53,2
Konji	53,8	41,5	61,3	49,1	71,7	54,6
Logotipi Apple	69,9	47,4	77,6	56,5	86,1	80,4
Steklenice	46,5	29,1	46,3	33,0	66,7	56,5
Skodelice	39,1	26,9	44,3	35,3	76,2	64,8



Slika 5.18: Primeri manifestacij značilnic Ferrari in sod. na testnih slikah. Za vsako kategorijo je prikazan po en primer nastopa značilnice, ki jo postopek AdaBoost izbere kot prvo (tabela 5.13).

Tabela 5.16: Prvih deset značilnic Ferrari in sod., izbranih po metodi notranje optimizacije.

Kategorija	Prvih 10 izbranih značilnic
Avtomobili	
Konji	
Logotipi Apple	
Steklenice	
Skodelice	

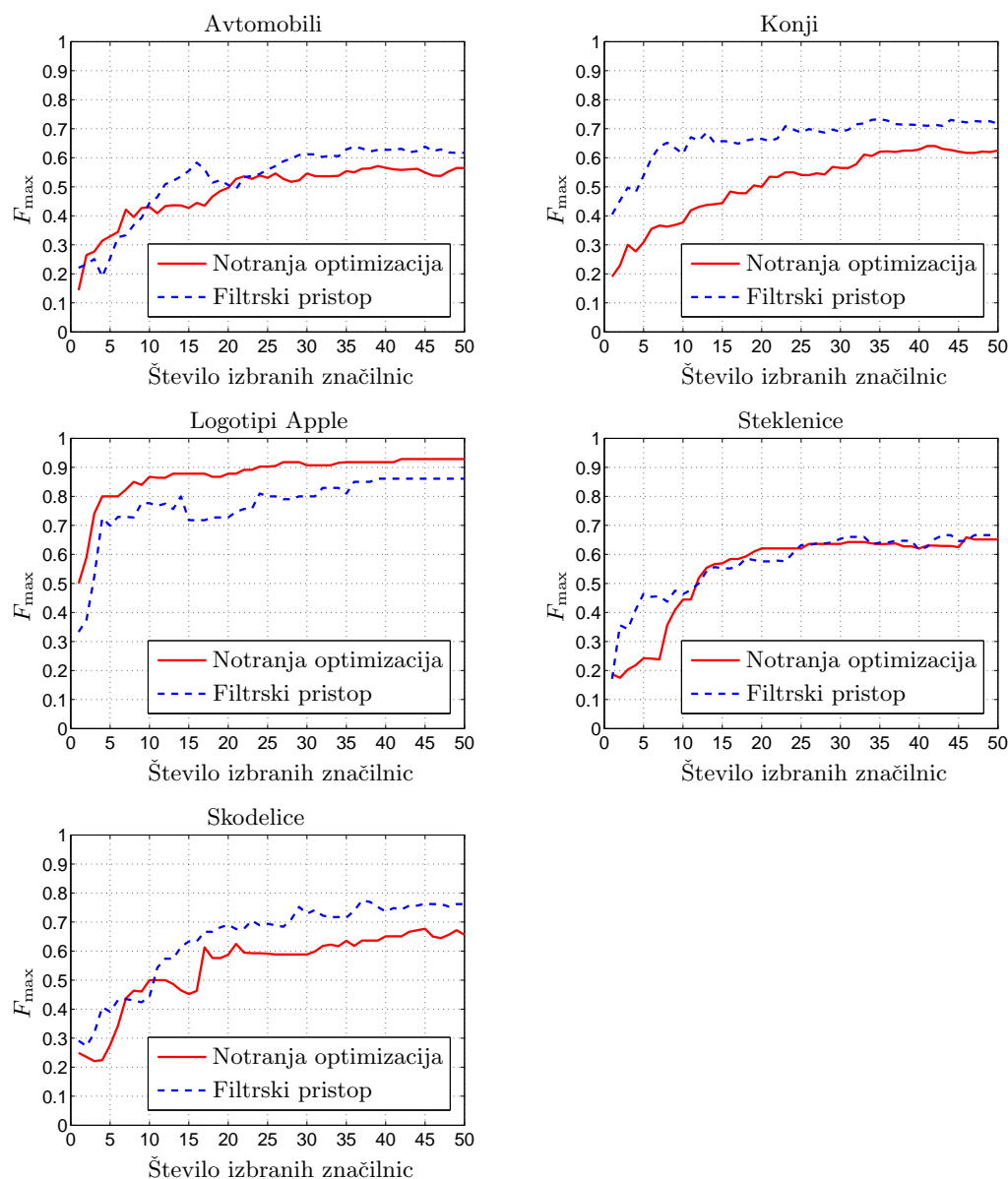
osnovi filtrske izbire tovrstnih značilnic. Krivulje odvisnosti $F_{\max}(T)$ so prikazane na sliki 5.19, podatki F_{\max} in EER za posamezna števila izbranih značilnic pa so zbrani v ustreznih odsekih tabel 5.17 in 5.18.

Presenetljivo dobro pa se notranja optimizacija obnese pri kategoriji logotipov Apple. Krivulja $F_{\max}(T)$ je primerljiva celo s krivuljo, pridobljeno na osnovi filtrske izbire značilnic Fidler in sod. Sicer drži, da je testna zbirka logotipov Apple razmeroma enostavna, kljub temu pa omenjena „anomalija“ namiguje na morebiten neizkoriščen potencial značilnic Ferrari in sod. za detekcijo objektov vizualnih kategorij.

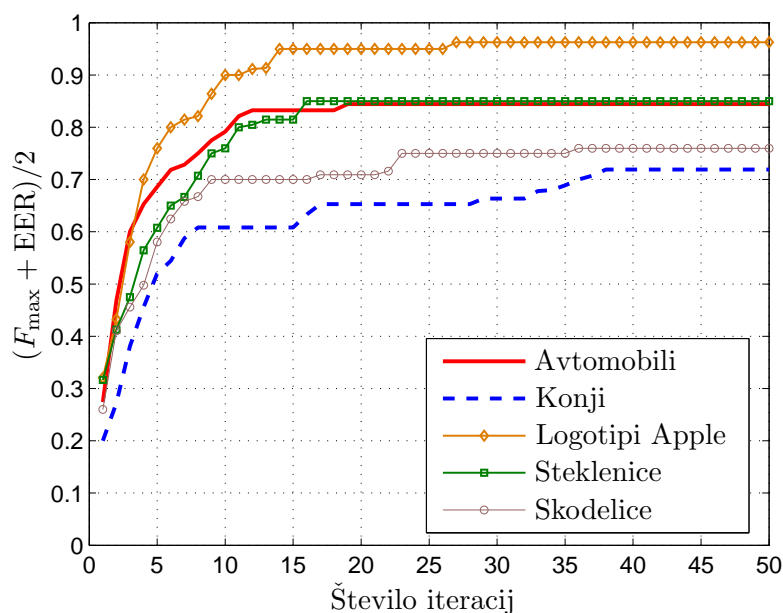
Tabela 5.16 prikazuje za vsako kategorijo prvih 10 značilnic Ferrari in sod., izbranih po metodi na osnovi notranje optimizacije. Diagram na sliki 5.20 prikazuje, kako se vrednost ocene $(F_{\max} + \text{EER})/2$ na učni množici \mathcal{I}_i spreminja v odvisnosti od števila požrešno izbranih značilnic.

5.7 Povzetek poglavja

V tem poglavju smo prikazali rezultate preizkusov vseh štirih kombinacij predstavljenih metod za pridobivanje značilnic in metod za izbiro značilnic. Vsako kombinacijo smo preizkusili na (istih) petih testnih slikovnih množicah. Vsaka testna množica opredeljuje neko vizualno kategorijo in posledično samostojno detekcijsko nalogo. Za vsako kombinacijo izbirne metode, sestava značilnic in testne množice smo pripravili ustrezno zbirko učnih slikovnih množic, ki jih za svoje delovanje potrebujejo postopek za pridobivanje značilnic, postopek za gradnjo detekcijskega



Slika 5.19: Odvisnost F_{\max} od števila izbranih značilnic Ferrari in sod. Na sliki primerjamo izbiro na osnovi notranje optimizacije (polne rdeče krivulje) z izbiro na podlagi postopka AdaBoost (črtkane modre krivulje).



Slika 5.20: Spreminjanje kriterija $(F_{\max} + EER)/2$ na množici \mathcal{I}_i pri požrešnem izbiranju značilnic Ferrari in sod. v sklopu notranjeoptimizacijske izbirne metode.

Tabela 5.17: Odvisnost detekcijske točnosti, podane kot $100 \times F_{\max}$, od sestava značilnic, izbirne metode in števila izbranih značilnic, ki znaša 5, 10 ali 50. Stolpec $|\mathcal{Z}|$ podaja število pridobljenih značilnic, stolpec „Vse“ pa navaja detekcijske rezultate, dobljene z uporabo vseh pridobljenih značilnic.

Značilnice Fidler in sod.								
Kategorija	$ \mathcal{Z} $	Vse	AdaBoost			Notranja opt.		
			5	10	50	5	10	50
Avtomobili	920	80,6	92,3	90,8	92,5	71,5	63,9	53,9
Konji	896	71,6	69,4	81,7	83,0	63,7	68,3	74,9
Logotipi Apple	592	84,3	83,5	84,1	89,2	86,4	84,6	89,4
Steklenice	896	68,6	59,3	61,4	72,2	39,7	52,7	63,8
Skodelice	912	74,4	64,6	75,6	83,5	43,9	48,9	67,8
Značilnice Ferrari in sod.								
Kategorija	$ \mathcal{Z} $	Vse	AdaBoost			Notranja opt.		
			5	10	50	5	10	50
Avtomobili	386	64,2	25,3	44,4	61,7	32,9	43,1	56,5
Konji	398	64,7	53,8	61,3	71,7	30,9	37,7	62,5
Logotipi Apple	168	90,2	69,9	77,6	86,1	80,0	86,7	92,9
Steklenice	422	77,7	46,5	46,3	66,7	24,2	44,4	65,2
Skodelice	314	82,8	39,1	44,3	76,2	27,6	50,0	65,6

Tabela 5.18: Ta tabela je povsem analogna tabeli 5.17, le da so rezultati podani kot vrednosti EER namesto kot F_{max} .

Značilnice Fidler in sod.								
Kategorija	$ \mathcal{Z} $	Vse	AdaBoost			Notranja opt.		
			5	10	50	5	10	50
Avtomobili	920	79,9	91,4	89,9	91,4	69,8	61,9	53,2
Konji	896	70,0	69,4	81,1	82,2	61,7	66,7	74,4
Logotipi Apple	592	79,5	81,8	84,1	84,1	81,8	79,5	86,4
Steklenice	896	65,5	58,2	60,0	70,9	38,2	52,7	61,8
Skodelice	912	69,7	62,1	71,2	81,8	40,9	48,5	65,1
Značilnice Ferrari in sod.								
Kategorija	$ \mathcal{Z} $	Vse	AdaBoost			Notranja opt.		
			5	10	50	5	10	50
Avtomobili	386	62,6	22,3	43,2	60,4	22,3	41,0	54,7
Konji	398	62,8	50,6	60,6	70,0	28,9	35,6	60,0
Logotipi Apple	168	84,1	68,2	75,0	81,8	77,3	81,8	90,9
Steklenice	422	76,4	43,6	45,5	63,6	18,2	41,8	61,8
Skodelice	314	80,3	37,9	42,4	71,2	25,8	45,5	63,6

modela za posamezne pridobljene značilnice in postopek za izbiro značilnic. Pri pripravi učnih množic smo upoštevali posebnosti, ki veljajo za posamezne izbirne metode oziroma sestave značilnic.

Pokazali smo, da je že z zelo majhnim številom preprostih značilnic možno doseči precej visoko detekcijsko točnost. Ta trditev še posebej velja za kombinacijo filtrske izbirne metode, značilnic Fidler in sod. in testne zbirke avtomobilov UIUC, kjer smo že s štirimi izbranimi značilnicami dosegli rezultat, ki je povsem primerljiv z rezultatom, ki ga za zbirko UIUC navajajo Mutch in Lowe (2006). Filtrska izbira značilnic Fidler in sod. omogoča doseg ugodnih detekcijskih rezultatov tudi pri drugih testnih zbirkah.

Izbira na osnovi notranje optimizacije je pripeljala do slabših rezultatov kot izbira po filtrski metodi. Izkazalo se je namreč, da celo požrešni postopek preiskovanja prostora stanj terja preveč časa, da bi bilo možno izvesti notranjeoptimizacijsko izbiro na dovolj veliki učni slikovni množici. Zaradi tega smo se morali zadovoljiti z zelo omejeno učno množico, ki pa ne omogoča dovolj natančnega razlikovanja med kandidatnimi množicami značilnic. Množica značilnic, ki vodi do visoke detekcijske točnosti na učni množici, lahko na neodvisni testni množici povsem odpove. To je tem bolj verjetno, čim manjša in čim manj raznolika je učna slikovna množica, na podlagi katere se izvaja izbira značilnic.

V okviru filtrske izbire značilnic smo primerjali izbirni postopek AdaBoost in izbiro po padajočih vrednostih ocenjevalne funkcije. Empirično smo potrdili, da postopek AdaBoost vodi do občutno boljših detekcijskih rezultatov. Razlog je v

medsebojni odvisnosti značilnic, ki je izbira značilnic po padajočih vrednostih ocenjevalne funkcije ne more upoštevati, postopek AdaBoost pa jo obravnava preko razmeroma preprostega mehanizma uteževanja učnih slik.

Uporaba sestava značilnic Fidler in sod. je v skoraj vseh primerih prinesla občutno boljše detekcijske rezultate kot uporaba sestava Ferrari in sod. Razlika je še posebej očitna pri kategoriji avtomobilov, saj z značilnicami Ferrari in sod. v nasprotju z značilnicami Fidler in sod. ni mogoče enostavno zajeti lokalne strukture avtomobilskega kolesa, ki predstavlja ključno oporno točko za zanesljivo detekcijo avtomobilov. Tudi pri drugih kategorijah se je sestav Fidler in sod. izkazal bolje od sestava Ferrari in sod. Nekaj verjetnih, četudi ne formalno dokazanih vzrokov za slabšo detekcijsko sposobnost sestava Ferrari in sod. v okviru implementiranega sistema smo navedli v razdelku 5.5.

Na osnovi dobljenih rezultatov bi bilo torej najbolj smiselno priporočiti uporabo predlagane filtrske izbirne metode v povezavi s sestavom značilnic Fidler in sod. Kot smo omenili, gre zasluge za uspeh te kombinacije pripisati tudi mehanizmu delovanja detekcijskega postopka. Razmeroma nizka stopnja specifičnosti izbranih značilnic Fidler in sod., ki je posledica njihove enostavnosti, se v kombinaciji z implementiranim detekcijskim postopkom povečini izkaže za manjše zlo kot nizka stopnja ponovljivosti, ki jo izkazujejo nekatere izbrane značilnice Ferrari in sod.

Izvedli smo še nekaj dodatnih preizkusov. Pokazali smo, da postopek MDL občutno pripomore k izboljšanju detekcijske točnosti. Ta trditev velja za vse kombinacije izbirnih metod in sestavov značilnic. Poleg tega smo pokazali, da je tudi pri uporabi značilnic Ferrari in sod. smiselno obravnavati testne slike v več velikostnih različicah, čeprav to zaradi velikostno invariantne opredelitve značilnic Ferrari in sod. formalno ne bi bilo potrebno.

Zaključek

6.1 Pregledni povzetek magistrske naloge

V magistrski nalogi smo preučili problem izbire značilnic za detekcijo objektov dane vizualne kategorije na slikah. Izbira značilnic lahko izboljša detekcijsko zmogljivost v dveh pogledih. Manjše število značilnic, ki jih mora obravnavati detekcijski postopek, neposredno vodi do manjše časovne zahtevnosti. Poleg tega pa se pogosto izkaže, da izhodiščna množica značilnic, pridobljena iz neke slikovne množice, vsebuje določeno število za detekcijo povsem nekoristnih ali celo škodljivih značilnic. V takšnih primerih izbira značilnic poveča tudi točnost, ne zgolj računsko učinkovitost detekcije objektov na testnih slikah.

Implementirali in preizkusili smo dve izbirni metodi, filtrsko in notranjeoptimizacijsko. V okviru filtrske metode smo definirali ocenjevalno funkcijo, ki vrednoti koristnost značilnic za detekcijo objektov dane kategorije na učnih slikah. Intuitivni pojem koristnosti ocenjevanje značilnice smo formalizirali z uvedbo klasifikatorja, ki razvršča učne slike na podlagi rezultatov detekcijskega postopka, omejenega na uporabo ocenjevanje značilnice oziroma njenih manifestacij na učnih slikah. Ocena značilnice je definirana kot točnost klasifikatorja, povezanega z ocenjevano značilnico.

Ocenjevalno funkcijo smo zasnovali tako, da jo je mogoče enostavno vgraditi v izbirno ogrodje AdaBoost. Funkcijo je sicer možno uporabiti tudi samostojno — v tem primeru se značilnice izbirajo preprosto po padajočih ocenah — vendar postopek AdaBoost (implicitno) upošteva odvisnosti med značilnicami, zato pri danem ciljnem številu značilnic T vodi do izbirne množice, ki v splošnem izkazuje bistveno manjšo stopnjo redundance kot nabor T najbolj ocenjenih značilnic.

Notranjeoptimizacijska izbirna metoda sistematično izvaja detekcijski postopek na učnih slikah za različne kandidatne množice izbranih značilnic. Rezultat vsakega posameznega izvajanja je detekcijska točnost za ustrezno kandidatno izbirno množico. Namen izbirnega postopka je poiskati izbirno množico, ki vodi do najvišje detekcijske točnosti pri danem ciljnem številu značilnic. Ker je število vseh možnih kandidatnih izbirnih množic preveliko za popoln prebor, je smiselno vzpostaviti bijektivno preslikavo med kandidatnimi množicami in vozlišči v grafu

prostora stanj, nato pa na dobljenem grafu izvesti kak suboptimalen iskalni postopek. V praksi se zaradi velike časovne zahtevnosti notranjeoptimizacijskega izbiranja običajno uporablja kar požrešno iskanje.

Predstavljena izbirna pristopa je mogoče uporabiti v povezavi s širokim razponom detekcijskih metod. V sistemu, ki smo ga izdelali, smo implementirali in nekoliko prilagodili detekcijski pristop, ki so ga predlagali Leibe in sod. (2004). Ta pristop, ki temelji na paradigmi implicitnega modela oblike, se je izkazal že v marsikaterem detekcijskem problemu, zato smo se ga odločili preizkusiti tudi v magistrski nalogi.

Jedro detekcijskega postopka predstavlja glasovanje za hipoteze o središčih domnevnih objektov na osnovi položajev manifestacij posameznih značilnic. Fazi glasovanja sledi robustno iskanje detekcijskih hipotez. Izredno močno nadgradnjo osnovnega detekcijskega postopka predstavlja postopek za dušenje sekundarnih hipotez na osnovi načela MDL. Detekcijski postopek je mogoče razmeroma enostavno prilagoditi za primer, ko testne slike prikazujejo objekte pri poljubnih velikostih. Po eni od možnih rešitev se vsaka testna slika obravnava v okviru velikostne piramide, pri čemer se vsaka plast piramide obdela z enovelikostnim detekcijskim postopkom. Pri uporabi značilnic, definiranih na velikostno invarianten način, pa je bolj smiselno izvajati različico detekcijskega postopka, ki velikostno invariantnost značilnic upošteva že pri učenju detekcijskega modela.

Vsako od obeh izbirnih metod smo povezali z dvema različnima pristopoma za pridobivanje in uporabo sestava značilnic. Pristop, ki so ga zasnovali Fidler in sod. (2006), temelji na hierarhični kompozicijski predstavitvi značilnic. Vsaka neelementarna značilnica je predstavljena kot fleksibilna geometrijska kompozicija ene ali več enostavnejših značilnic. Postopek za učenje hierarhije značilnic na osnovi slikovne množice deluje po načelu združevanja statistično pomembnih kombinacij enostavnejših značilnic v kompleksnejše kompozicije. Zasnova sestava značilnic omogoča učinkovito in robustno reprezentacijo neznanih slik z naborom pridobljenih značilnic.

Pristop, ki so ga zasnovali Ferrari in sod. (2006a), temelji na povsem drugačni podmeni. Metoda za pridobivanje značilnic Ferrari in sod. izhaja iz robne reprezentacije učnih slik. V začetnem koraku se povezani robni elementi združujejo v približno ravne odseke, ti pa se v naslednjem koraku združujejo v skupine po k odsekov. Vsaka posamezna značilnica Ferrari in sod. je definirana kot gruča podobnih skupin k robnih odsekov, pridobljenih iz učnih slik. Gručenje omogoča razmeroma robustno reprezentacijo neznanih slik s podanimi značilnicami. Za razliko od značilnic Fidler in sod. so značilnice Ferrari in sod. predstavljene na velikostno invarianten način.

Predstavljene metode smo eksperimentalno ovrednotili v skladu s priporočili, ki so jih objavili Agarwal in sod. (2004) in Leibe (2004). Sledi povzetek nekaterih pomembnejših empiričnih izsledkov:

- Izbira značilnic omogoča zmanjšanje porabe časa v testni (detekcijski) fazi.
- Ustrezno izbrana množica značilnic pogosto omogoča doseg višje detekcijske

točnosti kot izhodiščna množica značilnic.

- Izbirni postopek AdaBoost v skoraj vseh primerih vrne kakovostnejšo množico značilnic kot postopek, ki izbira značilnice po padajočih statičnih ocenah.
- V nekaterih primerih je mogoče že z majhnim številom preprostih značilnic doseči visoko detekcijsko točnost. Ta trditev še posebej velja za značilnice Fidler in sod., izbrane po filtrski metodi.
- Zaradi načina delovanja detekcijskega postopka je koristno, če izbrane značilnice ponovljivo nastopajo na vizualno podobnih lokalnih strukturah, medtem ko ima lastnost specifičnosti v splošnem manjšo, čeprav seveda nikakor ne zane-marljivo težo. To je eden od vzrokov za dejstvo, da je z značilnicami Fidler in sod. v okviru eksperimentalnih pogojev mogoče doseči višjo detekcijsko točnost kot z značilnicami Ferrari in sod.

6.2 Smeri nadaljnega razvoja

V odseku 1.2.2.3 smo pripomnili, da se notranjeoptimizacijske izbirne metode spričo svoje računske zahtevnosti pogosto uporabljajo le kot dopolnilo osnovnih (filtrskih) izbirnih metod. Peng in sod. (2005), na primer, s požrešnim iskanjem v smeri vzvratne eliminacije dodatno zmanjšajo moč (številnost) množice značilnic, ki jo je pripravila filtrska izbirna metoda.

Zamisli iz domene filtrskih in notranjeoptimizacijskih izbirnih metod je možno povezati tudi kako drugače. V izvedbi notranjeoptimizacijske metode, ki smo jo predstavili, je izbirni postopek „ovit“ okrog detekcijskega postopka. Vendar pa bi izbirni postopek bilo mogoče oviti tudi okrog ocenjevalne funkcije $u(Z; \mathbf{w})$. Prostor stanj bi bil v tem primeru definiran povsem enako kot v primeru ovitja okrog detekcijskega postopka: stanja bi ustrezala posameznim kandidatnim množicam značilnic. Tudi cilj notranjeoptimizacijske izbire bi ostal enak: poiskati najboljše ocenjeno stanje pri danem ciljnem številu značilnic. Oceno stanja $\{J(1), \dots, J(k)\}$ pa bi lahko opredelili kot $u(\{Z_{J(1)}, \dots, Z_{J(k)}\}; \mathbf{w}^{[1]})$, kjer je funkcija $u(\mathcal{F}; \mathbf{w})$ definirana kot razširitev funkcije $u(Z; \mathbf{w})$ na poljubno množico značilnic \mathcal{F} . Različica detekcijskega postopka, ki se uporablja pri določanju vrednosti ocenjevalne funkcije, v primeru določanja $u(Z; \mathbf{w})$ tvori detekcijske hipoteze na osnovi značilnice Z , v primeru določanja $u(\mathcal{F}; \mathbf{w})$ pa na osnovi množice značilnic \mathcal{F} .

Takšno kombinacijo filtrskega in notranjeoptimizacijskega pristopa smo tudi preizkusili. Uporabili smo enako izbirno strategijo kot v primeru običajnega notranjeoptimizacijskega postopka: požrešno izbiranje s pričetkom v stanju, ki ustreza prazni množici značilnic. Poslužili smo se istih učnih množic kot v primeru izbiranja značilnic po običajni filtrski metodi. Vendar pa se je izkazalo, da funkcija $u(\mathcal{F}; \mathbf{w}^{[1]})$ doseže „strop“ že po nekaj izbranih značilnicah, poleg tega v posameznih iteracijah veliko različnih izbir vodi do približno enake vrednosti $u(\mathcal{F}; \mathbf{w}^{[1]})$. Soočili smo se torej z enakimi težavami kot pri običajni notranjeoptimizacijski

metodi. Postopek bi bilo najbrž možno izboljšati le z bistvenim povečanjem moči in raznolikosti posameznih učnih množic.

Poleg različnih možnosti povezovanja predstavljenih izbirnih metod bi bilo izbirni sistem mogoče nadgraditi tudi v drugih smereh. V magistrski nalogi smo se ukvarjali le z gradnjo izbirne množice pri *podanem* ciljnem številu značilnic. Smiselna razširitev pristopa bi se zato posvetila tudi iskanju primerne števila izbranih značilnic. Sosledje filtrske in notranjeoptimizacijske metode, ki služi kot sredstvo za izbiro značilnic pri pristopu Peng in sod. (2005), vključuje tudi enostaven kriterij, ki določi najprimernejšo moč ciljne izbirne množice.

Magistrska naloga je obravnavala izbiro značilnic za detekcijo objektov vnaprej podane vizualne kategorije. Kot možna smer nadaljnjega dela se zato ponuja izbira značilnic za problem večkategorijske detekcije, ki poleg določitve položajev in velikosti posameznih objektov, ki so v neznanem številu upodobljeni na dani testni sliki, zahteva tudi določitev kategorijskih pripadnosti posameznih objektov. Medtem ko običajna, enokategorijska detekcija temelji na razločevanju podob objektov in ozadja, je cilj večkategorijske detekcije doseči čim zanesljivejše ločevanje med objekti posameznih kategorij. Torralba in sod. (2004) ter Opelt in sod. (2006b) sta primera metod, ki posegata na področje izbire značilnic za večkategorijsko detekcijo.

Seznam možnih izboljšav in razširitev pa s tem seveda še zdaleč ni izčrpan. Pot do zadovoljive rešitve osrednjih problemov, ki smo se jih dotaknili v magistrski nalogi, je namreč še dolga.

Literatura

- S. Agarwal, A. Awan, in D. Roth. Learning to detect objects in images via a sparse, part-based representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11):1475–1490, 2004.
- H. Bischof, A. Leonardis, in A. Selb. MDL principle for robust vector quantization. *Pattern Analysis and Applications*, 2(1):59–72, 1999.
- A. Blum in P. Langley. Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1–2):245–271, 1997.
- E. Borenstein in S. Ullman. Class-specific, top-down segmentation. V: *European Conference on Computer Vision*, strani 109–124, 2002.
- A.C. Bovik, M. Clark, in W.S. Geisler. Multichannel texture analysis using localized spatial filters. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 12(1):55–73, 1990.
- D. Comaniciu in P. Meer. Distribution free decomposition of multivariate data. *Pattern Analysis and Applications*, 2(1):22–30, 1999.
- T.M. Cover. The best two independent measurements are not the two best. *IEEE Transactions on Systems, Man, and Cybernetics*, 4(1):116–117, 1974.
- J.L. Crowley. *A Representation for Visual Information*. Doktorska dizertacija, Carnegie-Mellon University, 1981.
- G. Csurka, C. Dance, L. Fan, J. Williamowski, in C. Bray. Visual categorization with bags of keypoints. V: *ECCV'04 Workshop on Statistical Learning in Computer Vision*, strani 1–22, 2004.
- Gy. Dorkó. *Selection of discriminative regions and local descriptors for generic object class recognition*. Doktorska dizertacija, INP Grenoble, 2006.
- Gy. Dorkó in C. Schmid. Selection of scale-invariant parts for object class recognition. V: *International Conference on Computer Vision*, strani 634–640, 2003.

- R.O. Duda, P.E. Hart, in D.G. Stork. *Pattern Classification*. Wiley Interscience, 2 izdaja, 2000.
- B. Epshtein in S. Ullman. Feature hierarchies for object classification. V: *International Conference on Computer Vision*, strani 220–227, 2005.
- R. Fergus, P. Perona, in A. Zisserman. Object class recognition by unsupervised scale-invariant learning. V: *IEEE Conference on Computer Vision and Pattern Recognition*, strani 264–271, 2003.
- V. Ferrari, T. Tuytelaars, in L. Van Gool. Real-time affine region tracking and coplanar grouping. V: *IEEE Conference on Computer Vision and Pattern Recognition*, strani 226–233, 2001.
- V. Ferrari, L. Fevrier, F. Jurie, in C. Schmid. Groups of adjacent contour segments for object detection. Technical report, INRIA Rhône-Alpes, 2006a.
- V. Ferrari, T. Tuytelaars, in L. Van Gool. Object detection by contour segment networks. V: *European Conference on Computer Vision*, strani 14–28, 2006b.
- S. Fidler in A. Leonardis. Towards scalable representations of object categories: Learning a hierarchy of parts. V: *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.
- S. Fidler, G. Berginc, in A. Leonardis. Hierarchical statistical learning of generic parts of object structure. V: *IEEE Conference on Computer Vision and Pattern Recognition*, strani 182–189, 2006.
- F. Fleuret. Fast binary feature selection with conditional mutual information. *Journal of Machine Learning Research*, 5:1531–1555, 2004.
- Y. Freund in R. Schapire. A decision-theoretic generalization of online learning. *Computer and System Sciences*, 55(1):119–139, 1997.
- M. Fritz, B. Leibe, B. Caputo, in B. Schiele. Integrating representative and discriminant models for object category detection. V: *International Conference on Computer Vision*, strani 1363–1370, 2005.
- D.E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, 1989.
- H. Grabner in H. Bischof. On-line boosting and vision. V: *IEEE Conference on Computer Vision and Pattern Recognition*, strani 260–267, 2006.
- H. Grabner, P.M. Roth, in H. Bischof. Eigenboosting: Combining discriminative and generative information. V: *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

- A. Jakulin in I. Bratko. Testing the significance of attribute interactions. V: *International Conference on Machine Learning*, strani 409–416, 2004.
- G.H. John, R. Kohavi, in K. Pfleger. Irrelevant features and the subset selection problem. V: *International Conference on Machine Learning*, strani 121–129, 1994.
- F. Jurie in B. Triggs. Creating efficient codebooks for visual recognition. V: *International Conference on Computer Vision*, strani 604–610, 2005.
- K. Kira in L.A. Rendell. A practical approach to feature selection. V: *International Conference on Machine Learning*, strani 249–256, 1992.
- S. Kirkpatrick, C.D. Gelatt, in M.P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.
- J.J. Koenderink. The structure of images. *Biological Cybernetics*, 50(5):363–370, 1984.
- R. Kohavi in G.H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1–2):273–324, 1997.
- I. Kononenko. Estimating attributes: Analysis and extensions of RELIEF. V: *European Conference on Machine Learning*, strani 171–182, 1994.
- I. Kononenko. *Strojno učenje*. Založba FE in FRI, 2005.
- G. Lance in W. Williams. A general theory of classificatory sorting strategies: II. Clustering systems. *Computer Journal*, 10:271–277, 1967.
- B. Leibe. *Interleaved object categorization and segmentation*. Doktorska dizertacija, ETH Zürich, 2004.
- B. Leibe, A. Leonardis, in B. Schiele. Combined object categorization and segmentation with an implicit shape model. V: *ECCV'04 Workshop on Statistical Learning in Computer Vision*, strani 17–32, 2004.
- A. Leonardis, A. Gupta, in R. Bajcsy. Segmentation of range images as the search for geometric parametric models. *International Journal of Computer Vision*, 14(3):253–277, 1995.
- A. Leonardis, H. Bischof, in J. Maver. Multiple eigenspaces. *Pattern Recognition*, 35(11):2613–2627, 2002.
- F. Li, J. Košecká, in H. Wechsler. Strangeness based feature selection for part based recognition. V: *IEEE Conference on Computer Vision and Pattern Recognition Workshop*, stran 22, 2006.
- Y. Li in W. Ito. Shape parameter optimization for AdaBoosted active shape model. V: *International Conference on Computer Vision*, strani 251–258, 2005.

- T. Lindeberg. Scale-space theory: A basic tool for analysing structures at different scales. *Journal of Applied Statistics*, 21(2):224–270, 1994.
- D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- D. Martin, C. Fowlkes, in J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(5):530–549, 2004.
- K. Mikolajczyk, B. Leibe, in B. Schiele. Multiple object class detection with a generative model. V: *IEEE Conference on Computer Vision and Pattern Recognition*, strani 26–33, 2006.
- T.M. Mitchell. *Machine Learning*. McGraw Hill, 1997.
- J. Mutch in D.G. Lowe. Multiclass object recognition with sparse, localized features. V: *IEEE Conference on Computer Vision and Pattern Recognition*, strani 11–18, 2006.
- D. Omerčević, O. Drbohlav, in A. Leonardis. High-dimensional feature matching: Employing the concept of meaningful nearest neighbors. V: *International Conference on Computer Vision*, 2007.
- A. Opelt, A. Pinz, M. Fussenegger, in P. Auer. Generic object recognition with boosting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3):416–431, 2006a.
- A. Opelt, A. Pinz, in A. Zisserman. Incremental learning of object detectors using a visual shape alphabet. V: *IEEE Conference on Computer Vision and Pattern Recognition*, 2006b.
- A. Opelt, A. Pinz, in A. Zisserman. A boundary-fragment-model for object detection. V: *European Conference on Computer Vision*, strani 575–588, 2006c.
- H. Peng, F. Long, in C. Ding. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, 2005.
- J. Shotton, A. Blake, in R. Cipolla. Contour-based learning for object detection. V: *International Conference on Computer Vision*, strani 503–510, 2005.
- D. Skočaj, G. Berginc, B. Ridge, A. Štimec, M. Jogan, O. Vanek, A. Leonardis, M. Hutter, in N. Hawes. A system for continuous learning of visual concepts. V: *International Conference on Computer Vision Systems*, 2007.
- A. Thomas, V. Ferrari, B. Leibe, T. Tuytelaars, B. Schiele, in L. Van Gool. Towards multi-view object class detection. V: *IEEE Conference on Computer Vision and Pattern Recognition*, strani 1589–1596, 2006.

-
- A. Torralba, K.P. Murphy, in W.T. Freeman. Sharing features: efficient boosting procedures for multiclass object detection. V: *IEEE Conference on Computer Vision and Pattern Recognition*, strani 762–769, 2004.
- P. Viola in M.J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.
- W. Zhang, B. Yu, G.J. Zelinsky, in D. Samaras. Object class recognition using multiple layer boosting with heterogeneous features. V: *IEEE Conference on Computer Vision and Pattern Recognition*, strani 323–330, 2005.

Izjava o samostojnosti dela

Izjavljam, da sem magistrsko delo izdelal samostojno pod vodstvom mentorja prof. dr. Aleša Leonardisa. Izkazano pomoč drugih sodelavcev sem v celoti navedel v zahvali.

Luka Fürst